

VISIT...

LANZAROTE
Caliente.COM

KIRK-OTHMER

**ENCYCLOPEDIA OF
CHEMICAL
TECHNOLOGY**

EDITING - by **WATCHER**

VOLUME 14

FOURTH EDITION

**IMAGING TECHNOLOGY TO
LANTHANIDES**

CONTENT INDEX (vol 14)

(with hyperlinks)

Edited by <http://www.watcherworld.narod.ru/>

IMAGING TECHNOLOGY	1
IMINES, CYCLIC	2
IMMUNOASSAY	21
IMMUNOTHERAPEUTIC AGENTS	32
INCINERATORS	44
INCLUSION COMPOUNDS	61
INDIUM AND INDIUM COMPOUNDS	79
INDOLE	83
INDUSTRIAL ANTIMICROBIAL AGENTS	91
INDUSTRIAL HYGIENE	103
INFORMATION RETRIEVAL	112
INFORMATION STORAGE MATERIALS	138
- OPTICAL	138
- MAGNETIC	170
INFRARED TECHNOLOGY AND RAMAN SPECTROSCOPY	191
- INFRARED TECHNOLOGY	191
- RAMAN SPECTROSCOPY	208
INITIATORS	219
- FREE-RADICAL INITIATORS	219
- ANIONIC INITIATORS	236
- CATIONIC INITIATORS	244
INKS	247
INORGANIC HIGH POLYMERS	256
INSECT CONTROL TECHNOLOGY	267
INSULATION, ACOUSTIC	311
INSULATION, ELECTRIC	322
INSULATION, THERMAL	331
INSULIN AND OTHER ANTIDIABETIC AGENTS	338
INTEGRATED CIRCUITS	344
IODINE AND IODINE COMPOUNDS	358
ION EXCHANGE	371
ION IMPLANTATION	390
IONOMERS	404
IRON	411
IRON BY DIRECT REDUCTION	424
IRON COMPOUNDS	433
ISOCYANATES, ORGANIC	446
ISOPRENE	462
ITACONIC ACID AND DERIVATIVES	472
KETENES, KETENE DIMERS, AND RELATED SUBSTANCES	473
KETONES	485
KINETIC MEASUREMENTS	507
LABORATORY INFORMATION MANAGEMENT SYSTEMS	516
LAMINATED MATERIALS, GLASS	523
LAMINATED MATERIALS, PLASTIC	531
LANTHANIDES	539

ICE.

See REFRIGERATION AND REFRIGERANTS; WATER.

ICE COLORS.

See AZO DYES.

ICE CREAM.

See MILK PRODUCTS.

IDITOL.

See SUGAR ALCOHOLS.

IDOSE.

See CARBOHYDRATES; SUGARS.

ILANG-ILANG OIL.

See OILS, ESSENTIAL.

ILMENITE.

See TITANIUM.

IMAGING TECHNOLOGY

Imaging systems, consisting of specialty chemicals and techniques, are used to produce copies or photographic representations of macroscopic entities that can be seen by the human eye. Moreover, imaging systems are utilized to produce representations of what is outside the range of human vision.

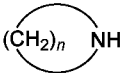
Visualization of the unseen, whether of the microscopic world of atoms and molecules, or of macroscopic materials hidden from view, has long been a goal of chemists, chemical engineers, and materials scientists wishing to comprehend the physical world. Increased capabilities in combined analytical methods (qv) such as spectroscopy (qv) (see also INFRARED TECHNOLOGY AND RAMAN SPECTROSCOPY; MAGNETIC SPIN RESONANCE), microscopy (qv), and spectrometry (see MASS SPECTROMETRY), aided by computer technology (qv) and advances in lasers (qv) as well as electronics and detection systems (see PHOTODETECTORS), have led to the production of images representative of individual atoms and molecules as well as those of aggregate surfaces and interfaces. The techniques of visualization are also used in medicine and in the fine arts.

An alphabetized list of *Encyclopedia* articles that are directly related to the various imaging technologies follows. This list is not meant to reflect every mention of or reference to imaging in the *Encyclopedia*, rather it is to serve as a guide to those articles where imaging is a primary concern.

Color photography	Photochemical technology
Color photography, instant	Photoconductive polymers
Electrically conductive polymers	Photography
Electrophotography	Printing processes
Fine art examination and conservation	Radioactive tracers
Information storage materials	Radiopaques
Lithographic resists	Surface and interface analysis
Medical imaging technology	Thermography
Nondestructive testing	X-ray technology

IMINES, CYCLIC

Ethyleneimine (aziridine, azacyclopropane) is the smallest cyclic imine consisting of a three-membered N-heterocyclic ring ($n = 2$):



This article describes ethyleneimine and the most important aziridine derivatives, not the higher homologues azetidine ($n = 3$), pyrrolidine ($n = 4$), or piperidine ($n = 5$). Reviews of these compounds are available (1–19). Unsubstituted ethyleneimine [151-56-4] is industrially the most important representative of the aziridine class. This substance was first prepared in the laboratory in 1888 (20,21), and independently confirmed in 1899 (22,23). Ethyleneimine was first synthesized industrially in 1938 by I. G. Farbenindustrie AG. After World War II ethyleneimine was produced by the Badische Anilin- und Soda-Fabrik Aktiengesellschaft (BASF AG) in Germany and by several companies in the United States (Chemirad Corporation, Dow Chemical Company, Cordova Chemical Company). In the 1960s, Interchemical Corporation and Union Carbide Corporation began large-scale production of ethyleneimine derivatives such as 2-methylaziridine and 1-(2-hydroxyethyl)aziridine. The BASF group is by far the largest manufacturer of ethyleneimine and has production plants in Germany and the United States. Another important producer is the Nippon Shokubai Company Ltd. of Japan.

Physical Properties

Ethyleneimine (EI) and its two most important derivatives, 2-methylaziridine [75-55-8] (propyleneimine) (PI) and 1-(2-hydroxyethyl)aziridine [1072-52-2] (HEA) are colorless liquids. They are miscible in all proportions with water and the majority of organic solvents. Ethyleneimine is not miscible with concentrated aqueous NaOH solutions (>17% by weight) (24). Ethyleneimine has an odor similar to ammonia and is detectable only at concentrations ≥ 2 ppm. The physical properties of ethyleneimine and the derivatives mentioned are given in Table 1. Thermodynamic data can be found in the literature (32).

Table 1. Physical Properties of Ethyleneimine and Derivatives

Property	EI	PI ^a	HEA ^b	Refs.
solidification point, °C	−74	−65		1,25
boiling point, °C	57	66	156	1,26,27
density, g/mol	0.837 ^c	0.8017 ^d	1.088	26–28
refractive index n_D	1.4130 ^c	1.4084 ^d	1.453 ^d	1,26,27
flashpoint, °C	−13	−10	67	25,26,28
ignition temperature, °C	322			29
ignition limits in air at 20°C, 101.3 kPa ^e				
lower explosion limit, vol %	3			29
upper explosion limit, vol %	55			29
viscosity at 25°C, mPa·s(=cP)	0.418	0.491		1,27
surface tension at 25°C, mN/m(=dyn/cm))	32.8			1
dielectric constant at 25°C	18.3			1
dipole moment, C·m ^f	7.8×10^{-30}			1
specific conductivity, (Ω·cm) ^{−1}	8×10^{-6}			1
heat of vaporization, kJ/mol ^g	34	33.2		1,26,27
heat capacity at 20°C, J/(g·K) ^g	2.48			
heat of combustion at 25°C, kJ/mol ^g	1.59×10^3			30
heat of formation, liquid, 25°C, kJ/mol ^g	92			30
heat of formation, gas, 25°C, kJ/mol ^g	127			1
heat of polymerization, kJ/kg ^g	2.3×10^3			1,26
heat of mixing with 80% H ₂ O, kJ/mol ^g	13.8			1
vapor pressure, kPa ^e (°C)	30.76 (27)	18.6		31,27

^a Propyleneimine.
^b Hydroxyethylaziridine.
^c At 20°C.
^d At 25°C.
^e To convert kPa to mm Hg, multiply by 7.5.
^f To convert C·m to debye, multiply by 2.99×10^{29} .
^g To convert J to cal, divide by 4.184.

Chemical Properties

Unlike the other structural isomers of C₂H₅N, *N*-methylenemethylamine (33,34), ethylideneimine (35), and vinylamine [593-67-9] (36,37) and the analogous phosphorus compound, phosphirane (38), ethyleneimine is stable at room temperature provided CO₂ is excluded from the air (39). Unexpectedly, ethyleneimine has the highest calculated relative heat of formation of the C₂H₅N isomers (40). Relative calculated heats of formation are ethylideneimine,

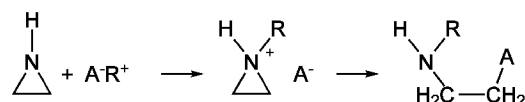
0.0; vinylamine, 26.9; *N*-methylenemethylamine, 32.2; and ethyleneimine, 85.3 kJ/mol (20.4 kcal/mol).

The special types of bonding in three-membered ethyleneimine rings (41–43) have been studied using microwave spectroscopy (44–47), electron diffraction (48), and photoelectron spectroscopy (49–51), and have occupied theoretical chemists up to the present day (52). These studies reveal that ethyleneimine has a distinctly shortened C—C bond of 0.148 nm (as compared to 0.154 nm in open-chain compounds) and a noticeably lengthened C—N bond of 0.149 nm (compared to 0.146 nm). Because of the high *s* character of the free electron pair on the nitrogen, ethyleneimine also shows a lower basicity ($pK_a = 7.98$) than noncyclic aliphatic amines such as dimethylamine ($pK_a = 10.7$) (53).

REACTIONS

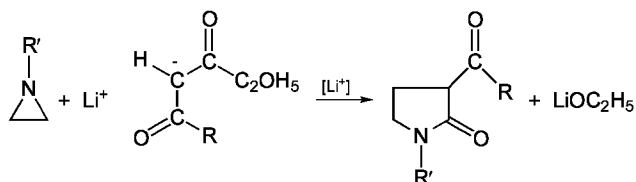
Depending on the experimental conditions used, the basicity or the ring strain can be the driving force in reactions involving ethyleneimine. With catalysis by Brønsted or Lewis acids, the aziridine ring can be opened by a large number of nucleophiles to give β -substituted ethylamines. In the absence of strong nucleophiles and at elevated temperatures, preparation of polyethyleneimines from aziridines is possible by acid-catalyzed reaction of the aziridine with itself. On the other hand, ethyleneimine and other aziridines substituted only on carbon show the typical reactions of a secondary amine, such as addition onto unsaturated systems, complex formation with metals, and reaction with halogen compounds. At low temperatures and alkaline pH the *N*-substituted aziridines are generally formed in these reactions. High temperatures and catalysis by acids or nucleophiles promote secondary reactions with opening of the three-membered ring, and these can be used for synthesis of heterocyclic compounds.

Nucleophilic Ring Opening. Opening of the ethyleneimine ring with acid catalysis can generally be accomplished by the formation of an intermediate aziridinium salt, with subsequent nucleophilic substitution on the carbon atom which loses the amino group. In the following, R represents a Lewis acid, usually H^+ ; A^- = the nucleophile.

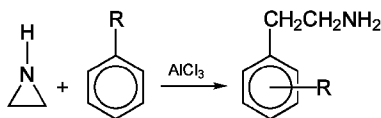


Because of the rapid ring opening by the nucleophile, aziridinium salts cannot usually be isolated. However, in a few cases it is possible to isolate such compounds (54), eg, at low temperatures, when the aziridinium salts are sparingly soluble or where there is steric hindrance to substitution. Stable ethyleneiminium salts can be prepared by reaction of ethyleneimine with acids not containing nucleophilic anions, for example HBF_4 (55).

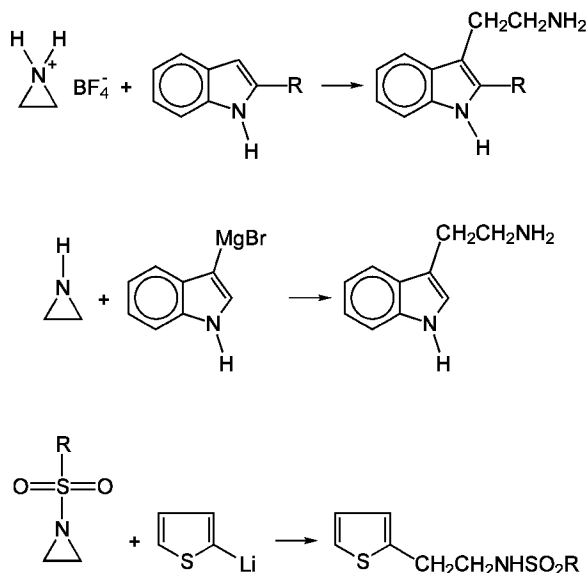
Reaction with Carbon Nucleophiles. Unactivated aziridines react with the lithium salts of malonates or β -keto esters in the presence of lithium salts to yield 3-substituted pyrrolidinones (56–59), where $R' = \text{alkyl and aryl}$, and R = alkoxy, alkyl, and aryl.



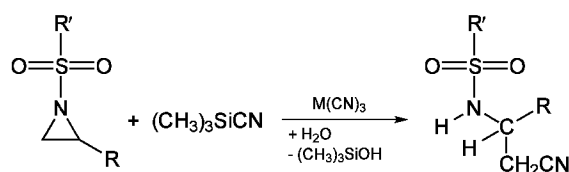
Carbocyclic aromatic compounds (R = H, C_2H_5 , Cl, OCH_3) can be aminoethylated in the presence of $AlCl_3$ (60–62).



N-sulfonated aziridines have also been used in Friedel-Crafts reactions (qv) (63). The successful C-alkylation of the heteroaromatic compounds indole (qv) [120-72-9] (64–66) and thiophene [110-02-1] (67) with aziridines has also been reported:



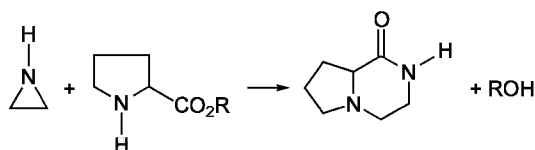
Direct reaction of hydrocyanic acid with ethyleneimine does not yield the desired β -aminopropionitrile (68). However, ring opening of tosylated aziridines to give the corresponding tosylated β -aminopropionitriles is possible using trimethylsilyl cyanide [7677-24-9] with lanthanoid tricyanide catalysis (69,70).



where M = Yb, Y, Ce; R = C₆H₅, C₄H₉, C₂H₄SCH₃, CH₂C₆H₅; R' = C₆H₄CH₃

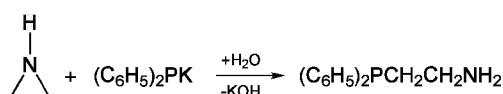
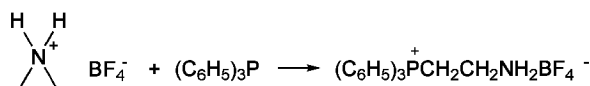
Reaction with Nitrogen Nucleophiles. The acid-catalyzed reaction of primary, secondary, and tertiary amines with ethyleneimine yields asymmetrically substituted ethylenediamines (71). Steric effects dominate basicity in the relative reactivity of various amines in the ring-opening reaction with ethyleneimine (72). The use of carbon dioxide as catalyst in the aminoethylation of aliphatic amines, for which a patent application has been filed (73), has two advantages. First, the corrosive salts produced when mineral acids are used as catalysts (74,75) are no longer formed, and second, the reaction proceeds with good yields under atmospheric pressure.

Aluminum chloride [7446-70-0] is a useful catalyst in the reaction of aromatic amines with ethyleneimine (76). Solid catalysts promote the reaction of ethyleneimine with ammonia in the gas phase to give ethylenediamine (77). Not only ammonia and amines, but also hydrazine [302-01-2] (78), hydrazoic acid [7782-79-8] (79–82), alkyl azidoformates (83), and acid amides, eg, sulfonamides (84) or 2,4-dioxypyrimidines (85), have been used as ring-opening reagents for ethyleneimine with nitrogen being the nucleophilic center (1). The 2-oxopiperazine skeleton has been synthesized from α -amino acid esters and ethyleneimine (86–89).

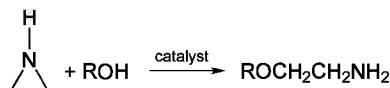


Ethyleneimine dimer has been synthesized using catalytic amounts of an alkali metal amide of ethyleneimine under alkaline conditions (89,90).

Reaction with Phosphorus Nucleophiles. The ethyleneimine ring can be opened using phosphines (91) or alkali metal phosphides (92):

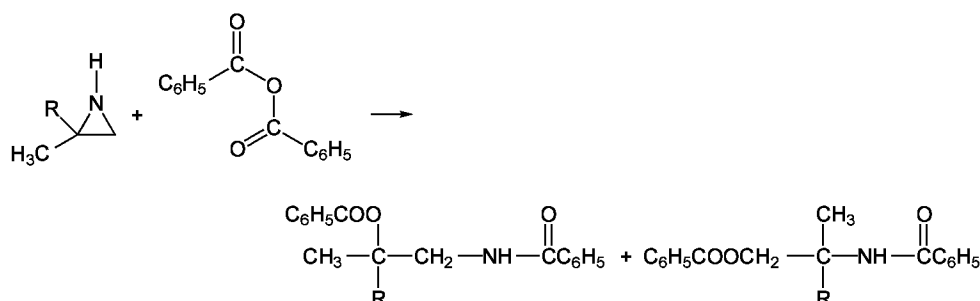


Reaction with Oxygen Nucleophiles. In the presence of strong acids, eg, H₂SO₄, HBF₄, or BF₃, aziridines react with alcohols to form β -amino ethers (93):

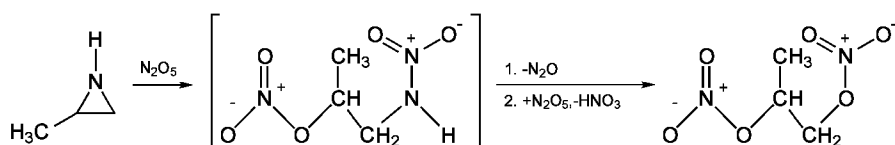


The reaction of a hydroperoxide with 2-methylaziridine [75-55-8] has been described (94). The reaction of ethyleneimine with phenols (95) and carboxylic acids (96,97) produces ethylamine ethers and esters, respectively. However, these reactions frequently yield product mixtures which contain polyaminoalkylated oxygen nucleophiles and polymers, in addition to the desired products (1). The selectivity of the reaction can often be improved by using less than the stoichiometric amount of the aziridine component (98,99).

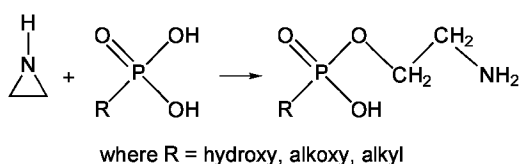
Rearrangement of the initial products to give amides or *N*-nitrophenylmonoethanolamines may occur as secondary reactions when aziridines react with carboxylic acids and phenols which have nitro groups in the ortho and para positions (1,100). In the reaction of aziridines with anhydrides, the attacking reagent simultaneously acts as oxygen nucleophile and as electrophilic agent. Ester amides are formed when carboxylic anhydrides react with aziridines (101–104).



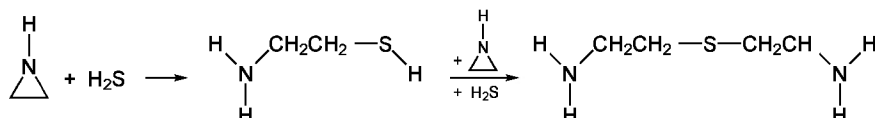
The anhydride of nitric acid, N₂O₅, reacts with 2-methylaziridine to give the dinitrate (105). In the case of *N*-substituted aziridines, the reaction stops at the stage of the nitramine nitrate prior to elimination of N₂O (106).



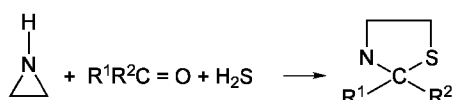
Phosphoric acid, monoalkyl phosphates, and phosphonic acids, but not dialkyl phosphates (107), can be aminoalkylated on the oxygen (108–110).



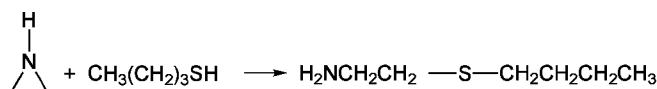
Reaction with Sulfur Nucleophiles. Because sulfur is highly nucleophilic, reactions of aziridines with sulfur nucleophiles generally proceed rapidly (111) and with good yields. The reaction of hydrogen sulfide [7783-06-4] with ethyleneimine yields cysteamine [60-23-1] (2-mercaptoethylamine) or bis(2-aminoethyl)sulfide [871-76-1] (2,112) depending on the molar ratio of the reactants. The use of NaHS for the synthesis of cysteamine has also been described (113).



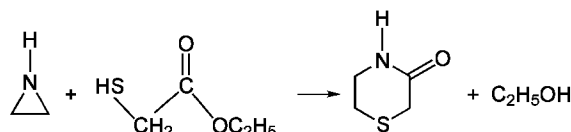
The reaction of hydrogen sulfide with aziridines in the presence of aldehydes or ketones provides a simple route to two-substituted thiazolidines (2,114–116).



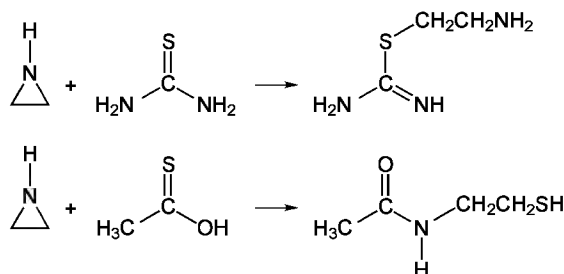
The reaction of aliphatic and aromatic mercaptans with aziridines yields thioethers (117–119).



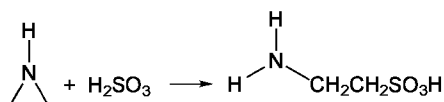
The di- or tetrahydro-1,4-thiazine skeleton is obtained if mercaptans which have a carbonyl group in the β -position react with ethyleneimine (120–124)



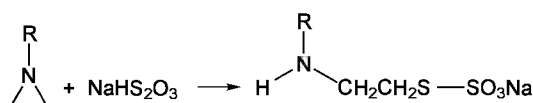
The reaction of thioethers with ethyleneimine in the presence of acid yields sulfonium compounds. The reaction is reversible under alkaline conditions (125). Compounds in which double-bonded sulfur can exist in tautomerism with a form having a free SH group, such as thiourea (126,127), thiocarboxylic acids (128), and thiophosphates (129), react to give aminoalkylated products. The β -aminoethyl thiocarboxylate rearranges to give the amide.



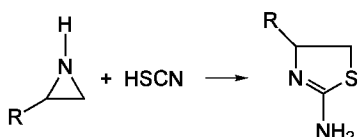
Ethyleneimine reacts rapidly with sulfurous acid to give taurine [107-35-7] in high yield, a reaction of importance not only for the preparation of this amino sulfonic acid but also for the decontamination of ethyleneimine solutions (130).



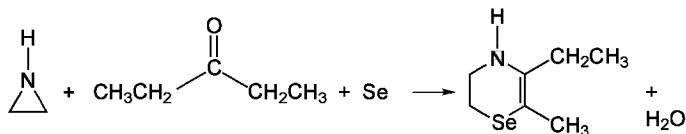
The reaction of aziridines (R = H, triazinyl) with thiosulfate yields *S*-alkyl thiosulfates (131,132), which are known as Bunte salts (133).



2-Amino-2-thiazolines are formed from thiocyanic acid [463-56-9] and aziridines where R = H or CH₃ (116,134).

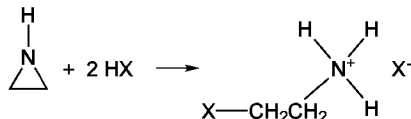


Reaction with Selenium Nucleophiles. The reactions of selenium nucleophiles are similar to those of the sulfur nucleophiles: selenophosphates can be aminoalkylated (135). A dihydroselenazine has been obtained by reaction of diethyl ketone, elementary selenium, and ethyleneimine (136).



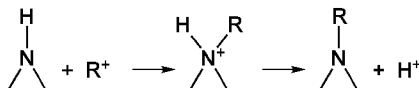
Selenosulfate reacts with ethyleneimine in the same way as thiosulfate to give 2-aminoethaneselenosulfuric acid. However, the reaction of ethyleneimine using selenous acid does not yield a stable product (137).

Reaction with Halogen Nucleophiles. Hydrochloric acid [7647-01-0], hydrobromic acid [10035-10-6], and hydroiodic acid [10034-85-2] react readily with ethyleneimine (3) to give the corresponding β -halogenoethylamines (20,21).

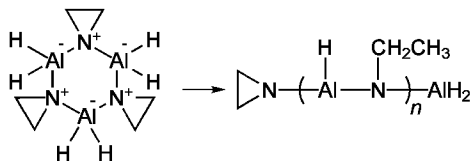


β -Fluoroethylamines are accessible by ring opening of aziridines with hydrogen fluoride in pyridine or with hydrogen fluoride and SOCl_2 at low temperatures in ether (138–141). β -Bromoalkylcyanamides have also been obtained by reaction of *N*-alkylated aziridines with cyanogen bromide (142). In this reaction nucleophilic ring opening by bromide and electrophilic attack by the CN group of cyanogen bromide on the aziridine nitrogen take place.

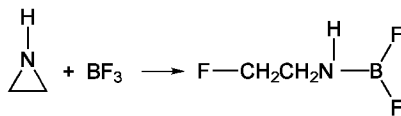
Electrophilic Reactions on the Aziridine Nitrogen. The generalized reaction of aziridines with an electrophile (R^+) is as follows.



Reactions with Electrophiles of Group IIIA (B,Al,Ga,In). Ethyleneimine forms cyclic trimers with hydrides or alkyls of elements of Group IIIA, with the liberation of hydrogen or hydrocarbons (143–146). In the case of diborane or trimethylboron, the initial adducts can be isolated (147–150). These aziridine boranes on the one hand undergo electrophilic reactions with opening of the aziridine ring, and on the other hand have the reducing properties of aziridine boranes (149) (see HYDROBORATION). Aziridinylalane trimer polymerizes on standing with reductive ring opening (145,151).

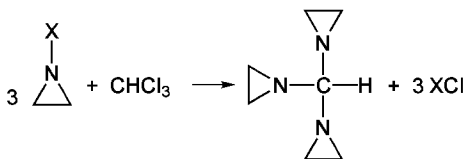


The reaction of 2-methylaziridine with boron trichloride [10294-34-5] leads to replacement of all three chlorides by aziridine rings to form tri(methylethyleneimine) boron [17862-61-2] (152). The reaction of boron trifluoride [7637-07-2] with ethyleneimine at -78°C proceeds via substitution and subsequent ring opening to yield *N*- β -fluoroethyl-*B*-difluoroborazene (153).

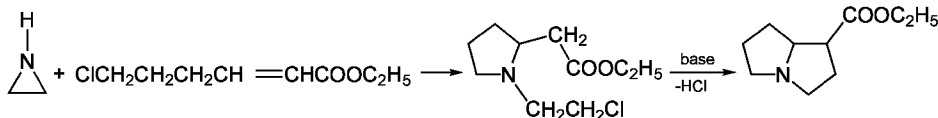


Reaction with Carbon Electrophiles. Halogenated carbon compounds, heterocyclic three-membered rings, and unsaturated carbon compounds containing a carbon–carbon or carbon–hetero atom multiple bond can act as electrophiles attacking the aziridine nitrogen via the carbon. The initial products are frequently capable of undergoing ring expansion to form larger heterocycles (7,8).

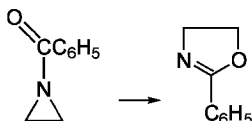
Aziridines ($\text{X} = \text{H}$) can be alkylated on the nitrogen, with retention of the three-membered ring, by reaction with aliphatic and aromatic halides in the presence of base (2,154). The reaction can also be carried out, in some cases with very good yields, under phase-transfer conditions using 30% NaOH and optionally an organic solvent (155). If the halides do not react readily, the alkali metal salts ($\text{X} = \text{Na}$) of the corresponding aziridine can be used (156–158) to form, for example, triethyleiminemethane [23974-29-0]..



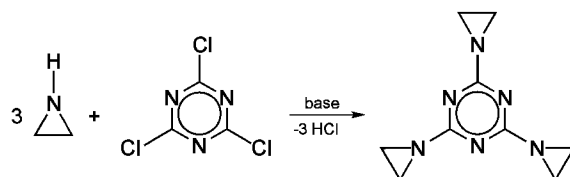
In the absence of a base, the aziridine ring can be quaternized and opened by the nucleophile. A pyrrolizidine synthesis, in which such a reaction proceeds intramolecularly followed by a Michael addition (159), is shown as follows:



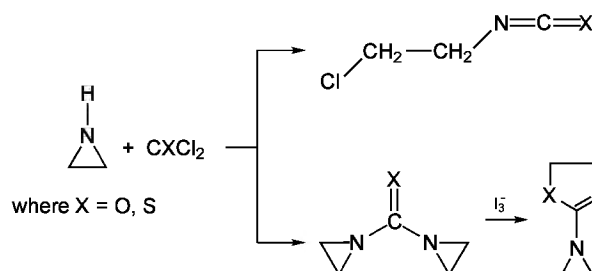
In the presence of a base, acid chlorides react readily with aziridines to give acylated aziridines (2,22,160–163). In the absence of a base, however, ring opening takes place and 2-chloroethylamides are obtained (2,164). Under suitable conditions acylated trialkylammonium salts of ethylenediamine can be prepared from acid chlorides, ethyleneimine, and tertiary amines (71). Acylated aziridines can be rearranged to 2-oxazolines by the action of heat, nucleophiles, or acids. The rearrangement of thioacylaziridines proceeds analogously (7,8,165–171).



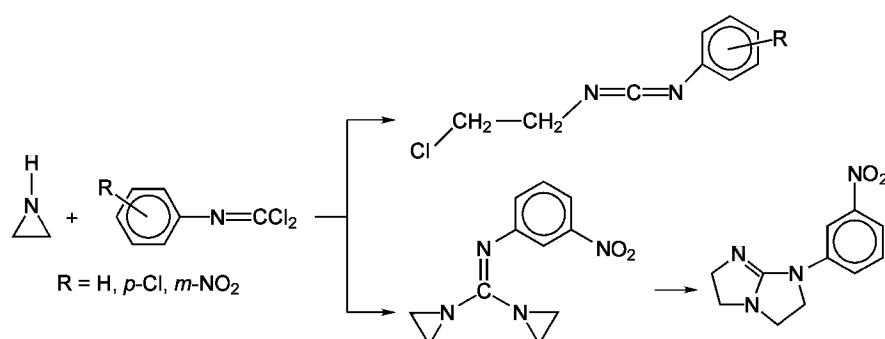
Reaction of cyanuric chloride [108-77-0] with ethyleneimine yields triethylenemelamine [51-18-3] (172).



The reaction of phosgene [75-44-5] or thiophosgene [463-71-8] with ethyleneimine yields either a bisaziridine compound (1,1'-carbonylbisaziridine [1192-75-2] or 1,1'-thiocarbonylbisaziridine [13163-23-0]) or a 2-chloroethyliso(thio)cyanate, depending on the reaction conditions. The former can be isomerized by catalysis with triiodide to give 1-aziridinyl-2-oxazoline [19587-77-0] or 1-aziridinyl-2-thiazoline [17205-48-0] (173–177).

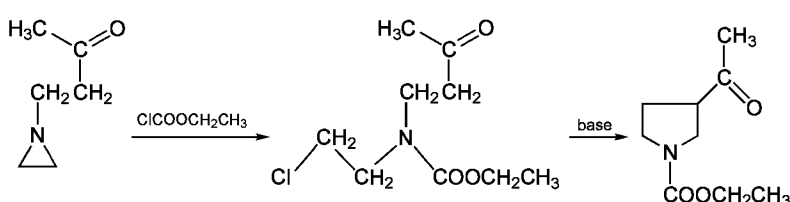


Bisaziridine compounds and *N*-(2-chloroethyl)carbodiimides have also been prepared using isocyanide dichlorides and ethyleneimine (178,179). The iodide-catalyzed rearrangement of the formerly mentioned compounds provides a method for preparing the tetrahydroimidazoimidazole system:

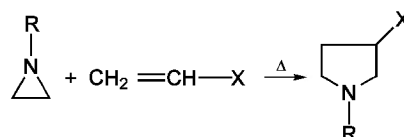


Ethyleneimine reacts with epoxides to form hydroxyalkylated products, eg, *N*-(β-hydroxyethylaziridine) [1072-52-2]. The epoxide component is frequently used in substoichiometric amount in order to prevent multiple alkoxylation (180–190). Ethyleneimine and episulfides react to give complex product mixtures, since the 1-(2-mercaptoethyl)aziridine produced initially can easily react further with both reactants (191,192).

Aziridines can add to carbon–carbon multiple bonds. Elevated temperature and alkali metal catalysis are required in the case of nonpolarized double bonds (193–195). On the other hand, the addition of aziridines onto the conjugated polarized double or triple bonds of α,β-unsaturated nitriles (196–199), ketones (197,200), esters (201–205), amides (197), sulfones (206–209), or quinones (210–212) in a Michael addition-type reaction frequently proceeds even at room temperature without a catalyst. The adducts obtained from the reaction of aziridines with α,β-unsaturated ketones, eg, 4-aziridinyl-2-butanone [503-12-8] from 3-buten-2-one, can be converted to 1,3-substituted pyrrolidines by subsequent ring opening with acyl chlorides and alkaline cyclization (213).

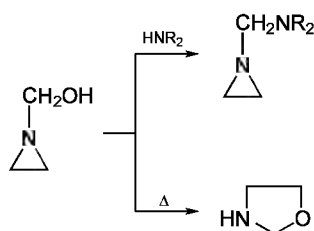


At temperatures >300°C, substituted pyrrolidines can be obtained by reaction of substituted aziridines ($R = \text{CH}_3, \text{C}_2\text{H}_5$) and conjugated olefins ($X = \text{CN}, \text{CO}_2\text{CH}_3, \text{CH}=\text{CH}_2$) with C–C cleavage in the three-membered ring (214–216).

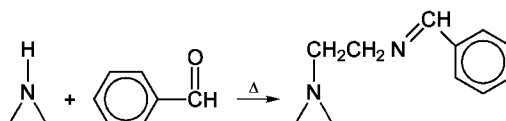


Reactions between vinyl ethers or vinyl acetate and ethyleneimine have not been satisfactory (198), but ethyleneimine does add onto the double bond of *N,N*-dimethylvinylamine to give 1-dimethylamino-1-ethyleneiminoethane [5498-98-6] (217).

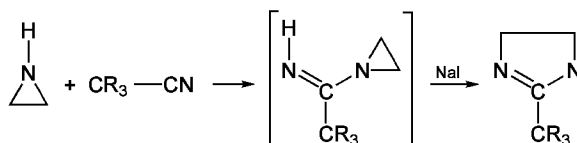
Aliphatic aldehydes and ketones react with aziridines to form relatively stable half aminals, eg, aziridine reacts with formaldehyde to form *N*-hydroxymethylaziridine [20276-43-1]. Half aminals can be converted to full aminals by reaction with a further secondary amine, isomerized to oxazolidines by the action of heat or used in a Mannich reaction for the ring aminomethylation of phenols, although this reaction gives only moderate yields (218–227).



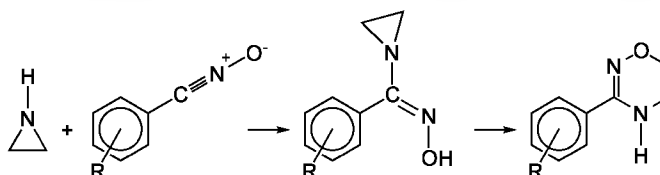
Schiff's bases of ethyleneimine dimer are obtained from the reaction of aromatic aldehydes, eg, benzaldehyde [100-52-7] or furfural [98-01-00], and ethyleneimine (228).



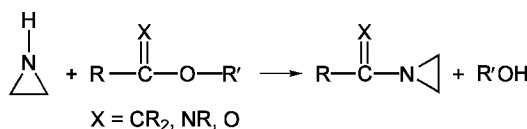
The reaction of ethyleneimine with nitriles in the presence of HBF_4 gives Δ^2 -imidazoles (229). If trichloroacetonitrile [545-06-2] ($\text{R} = \text{Cl}$) is used as the nitrile component, the intermediate amidine can be isolated (230).



1-Aroylaziridine oximes are accessible from aromatic nitrile oxides and aziridines and can rearrange to give the 1,2,4-oxadiazine derivatives (231–233).

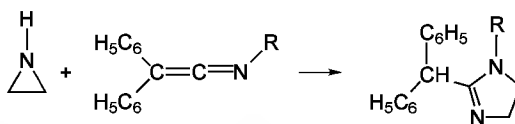
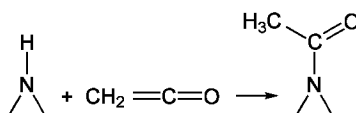


Acyl carbonates (234), alkoxyquinones (235) as vinylogous esters, imino ethers (236), and diketene (237) react with ethyleneimine to give the corresponding acylated ethyleneimines.

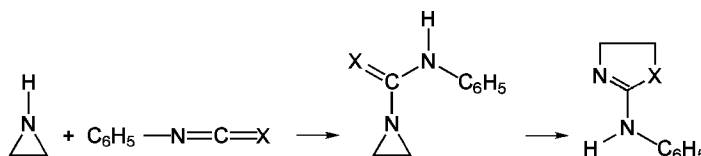


Reaction with esters of strong acids, such as formates or oxalates, yields the acyl derivatives of ethyleneimine dimer (238,239).

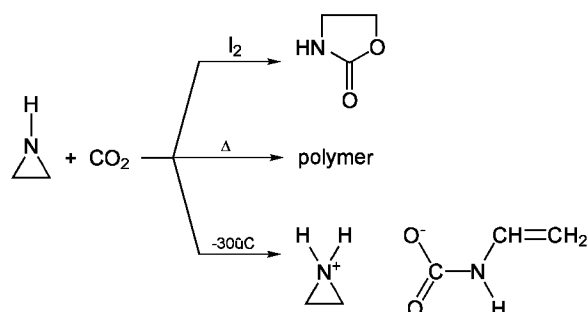
The reaction of heterocumulenes of the ketene type (ketenes, *N*-arylketeneimines, carbon suboxide) with aziridines leads to the formation of acylaziridines or imidoaziridines (240–242). In a reaction analogous to the ring expansion of acylaziridines, imidoaziridines rearrange in acids to give 2-substituted 2-imidazoles. These imidazoles are obtained directly in the reaction of ethyleneimine with keteneimines containing an aliphatic substituent on the nitrogen (242).



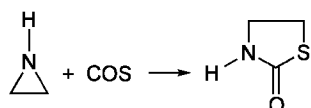
Carbodiimides ($\text{X} = \text{NC H}_2$), isocyanates ($\text{X} = \text{O}$), and isothiocyanates ($\text{X} = \text{S}$) also react with aziridines to give amidinoaziridines, carbamoylaziridines, and thiocarbamoylaziridines, respectively. As activated aziridine derivatives, these can rearrange to give derivatives of 2-amino-2-imidazoline, 2-amino-2-oxazoline, and 2-amino-2-thiazoline, respectively (243–250).



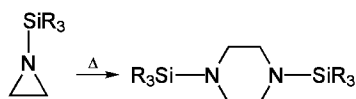
The iodine-catalyzed reaction of aziridines with carbon dioxide leads to 2-oxazolidinones (251). Because carbon dioxide effectively polymerizes ethyleneimine, only low yields are obtained when unsubstituted ethyleneimine reacts with CO_2 . However, direct insertion of carbon dioxide [124-38-9] into aziridines can be accomplished, with better yields, by ethoxycarbonylation of aziridines with subsequent elimination of ethylene under flash vacuum conditions (252). 1-Phenylaziridine [696-18-4] can react with CO_2 under antimony [7440-36-0] catalysis to give *N*-phenyl-2-oxazolidinone in good yields (253). At low temperatures and with the exclusion of atmospheric humidity, the reaction of ethyleneimine with carbon dioxide produces the unstable ethyleneiminium salt [31645-38-2] of *N*-vinylcarbamic acid (254,255).



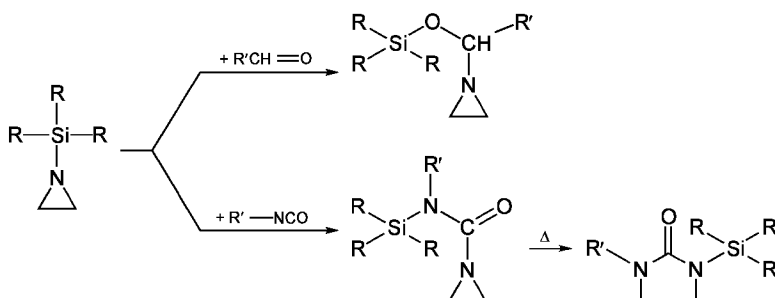
The reaction of ethyleneimine (and derivatives) with carbon oxysulfide yields 2-thiazolidinone [2682-49-7] (256,257). Carbon disulfide and ethyleneimine react to give 2-thiothiazolidine (258–260). Carbon diselenide reacts with aziridines to form 2-selenazolidineselenones (261).



Reaction with Further Electrophiles of Group IVA (Si, Ge, Sn). *N*-Silylated aziridines can be prepared from ethyleneimine by amination of chlorosilanes in the presence of an HCl acceptor, by dehydrocondensation with an organosilicon hydride or by cleavage of a silicon–carbon bond in 2-furyl-, 2-thienyl-, benzyl-, or allylsilanes in the presence of an alkali metal catalyst (262–266). *N*-Silylated aziridines can react with carboxylic anhydrides to give acylated aziridines, eg, *N*-acetylaziridine [460-07-1] in high yields (267). At high temperatures, *N*-silylaziridines can be dimerized to piperazines (268). Aldehydes can be inserted

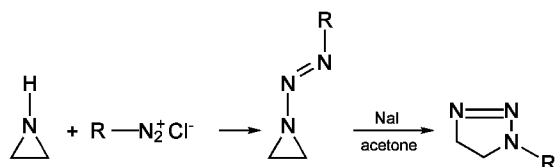
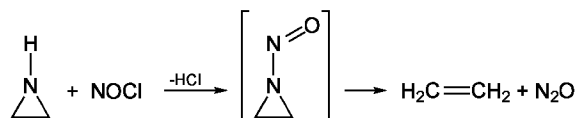


in the nitrogen–silicon linkage (269). The insertion of isocyanates with subsequent thermolysis and hydrolysis provides a method for the preparation of 1-alkyl-2-imidazolidinones (270).

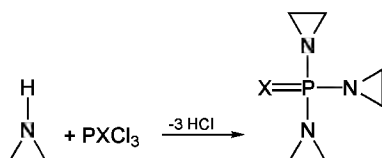


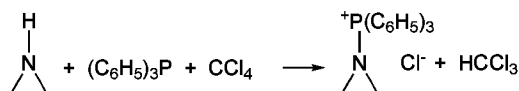
Alkylgermylaziridines of 2- and 4-valent germanium have also been described (271–273). Acetone can be inserted in the germanium–nitrogen linkage of triethylgermylaziridine (273) analogously to the above-mentioned reaction of silylaziridines with carbonyl compounds. *N*-Trimethylstannylaziridine [1481357-1] has been prepared by transamination of *N*-trimethylstannyldimethylamine with ethyleneimine (274), or by lithioamination of trimethyltin chloride or trimethyltin acetate [1118-14-5] with *N*-lithioethyleneimine (275). Carbon dioxide, carbon disulfide, phenyl isocyanate, acetone, or diethyl acetylenedicarboxylate can be inserted in the tin–nitrogen linkage (275).

Reaction with Electrophiles of Group VA (N, P, As). The reaction of aziridines with nitrosyl chloride (276,277) and other nitrosating reagents such as HNO₂, C₄H₉ONO, or NOBF₄ (18) proceeds via the thermally unstable *N*-nitrosoaziridines and leads to deamination. Coupling of aromatic diazonium salts with aziridines gives 1-aryldiazoaziridines, many of which are explosive. The 1-aryldiazoaziridines can be rearranged with sodium iodide in acetone to give *N*-substituted Δ²-1,2,3-triazolines (278,279).



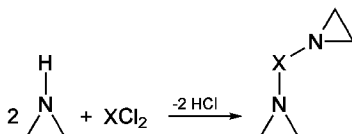
Suitable methods for linking a phosphorus–nitrogen bond to the aziridine ring are the aminolysis of halogenated phosphorus compounds (2,280–282), the transamination of phosphoramines with excess aziridine (283), the reaction with phosphites (284) and phosphoramidites (285) which have a free OH group, or the reaction of phosphines with aziridines and carbon tetrachloride (286).





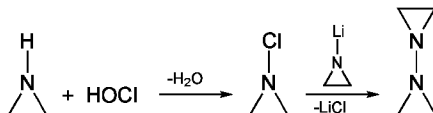
The reaction of monoalkylarsenic(III) halides and dialkylarsenic(III) halides with ethyleneimine has been described (287). This reaction proceeds successfully, unlike the reaction with AsCl_3 , which can lead to an explosion when purification by distillation is attempted.

Reaction with Sulfur Electrophiles. Bisaziridine compounds can be prepared from sulfur dichloride, thionyl chloride, or sulfuryl chloride and ethyleneimine (288). The products are, respectively, 1,1'-thiobisaziridine [2881-79-0] ($\text{X} = \text{S}$), 1,1'-dithiobisaziridine [1623-84-3] ($\text{X} = \text{S}-\text{S}$), 1,1'-sulfinylbisaziridine [1192-76-3] ($\text{X} = \text{SO}$), and 1,1'-sulfonylbisaziridine [931-92-0] ($\text{X} = \text{SO}_2$).



Diaziridinylsulfanes $\text{S}_x(\text{C}_2\text{H}_4\text{N})_2$, where $x = 1-5$, are also obtainable via diimidazolyisulfanes (289). These compounds tend to explode on distillation.

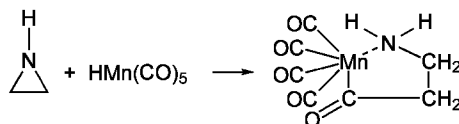
Reaction with Halogen Electrophiles. The synthesis of 1-haloaziridines, which are prone to explosion, has been carried out using hypohalites (290,291). 1-Chloroaziridine [10165-13-6] produced in this way reacts with 1-lithiated ethyleneimine to give 1,1'-diaziridine [4388-03-8]. Perchlorylaziridine [112405-46-6] has been prepared by reaction of ethyleneimine with dichlorine heptoxide at -20°C (292).



Reactions with Transition-Metal Compounds. The numerous published products of reactions of transition-metal compounds with aziridines can be divided into complexes in which the aziridine ring is intact, compounds formed by reaction of aziridine with the ligands of a complex, and complexes in which the aziridine molecule is fragmented (imido complexes).

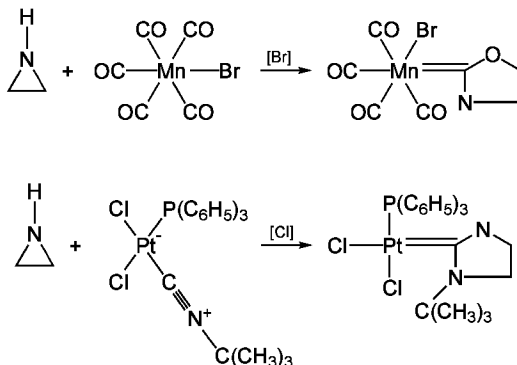
Aziridine Complexes in Which the Aziridine Ring is Intact. Central atoms which have been used for complexes in which the aziridine ring is intact include U (293), Ti (294–296), Zr (296), Cr (288–299), Mo (299), W (299), Mn (298–300), Fe (301), Co (297,298,300,302–304), Rh (305,306), Ni (298,300,303,307), Pd (297,298), Pt (297,298,308–310), Cu (298,300), Ag (300,311), Zn (300,311,312), Cd (300,311), and Hg (300). The metals Rh and Pt in low oxidation states form stable and inert aziridine complexes (313,314). Aziridine complexes of Zn and Ti are highly labile with respect to polymerization (313). The aziridine complexes of Co, Ni, and Cu fall between these two extremes. In complexes with central atoms of this type, the aziridine ligand can be dimerized to give 1-(2-aminoethyl)aziridine (299,311) or hexamerized to give 1,4-bis[2-(2-aminoethyl)aminoethyl]piperazine (315). Secondary reactions of the coordinated ethyleneimine with ring opening are also able to proceed with external nucleophiles, for example in nickel complexes with ammonium thiocyanate to give coordinated 2-amino-2-thiazoline (316).

β -Aminoacyl Complexes. Metallocarbonyl hydrides of Mo, W, Mn, and Co react with aziridines with ring opening of the aziridines and subsequent CO insertion to give β -aminoacyl complexes (317–319).

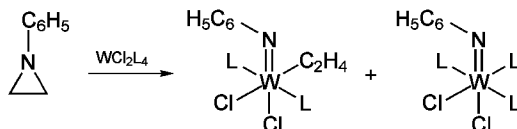


Oxidative cleavage of β -aminoacyl complexes can yield β -amino acid derivatives (320,321). The rhodium(I)-catalyzed carbonylation of substituted aziridines leads to β -lactams, presumably also via a β -aminoacyl-metal acyclic compound as intermediate. The substituent in the aziridine must have π or n electrons for coordination with the rhodium (322,323).

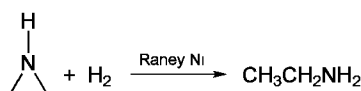
Cyclic Carbene Complexes. The reaction of aziridines with carbonyl, thiocarbonyl, or isonitrile ligands in Mn, Re, Fe, Ru, Pd, or Pt complexes leads to formation of cyclic carbene complexes (324–331).



Imido Complexes. The reaction of aziridines with tungsten(II) complexes leads to the formation of tungsten(IV) imido complexes (332):

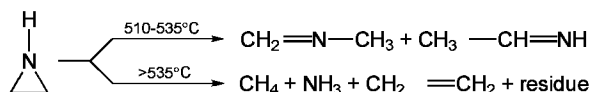


Reductive Ring Opening. Aziridines can be hydrogenated to ethylamines with catalysis by Raney nickel, palladium, or platinum (2,333–335). Lithium in ethylamine has also been used as reducing agent (336). Reductive ring opening of acylated aziridines has been performed using tributyltin hydride and methanol (337).



Oxidative Ring Opening. Many oxidizing reagents, such as peracids, ozone [10028-15-6], or FeI_2 , are suitable for oxidative deamination of aziridines to give olefins (18). On the other hand, oxidation of bicyclic 2,3-polymethyleneaziridines with lead tetraacetate leads to retention of the nitrogen in the molecule with the formation of ω -keto nitriles (338).

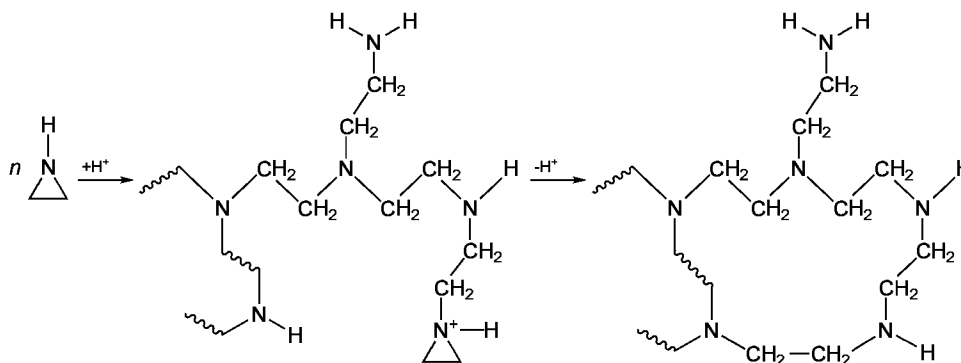
Thermal and Photochemical Reactions. Unsubstituted ethyleneimine has astonishing thermal stability. The reaction of ethyleneimine diluted with argon proceeds to give a mixture of unidentified compounds only at temperatures above 400°C (339). In a flow pyrolysis system under pressures of $<1.33\text{ kPa}$ ($<10\text{ mm Hg}$) on quartz wool, isomerization to give *N*-methylenemethylamine and ethyldeneimine was observed only in the temperature range $510\text{--}535^\circ\text{C}$. Higher temperatures result in fragmentation (340).



Irradiation of ethyleneimine (341,342) with light of short wavelength in the gas phase has been carried out directly and with sensitization (343–349). Photolysis products found were hydrogen, nitrogen, ethylene, ammonium, saturated hydrocarbons (methane, ethane, propane, *n*-butane), and the dimer of the ethyleneimino radical. The nature and the amount of the reaction products is highly dependent on the conditions used. For example, the photoproducts identified in a fast flow photoreactor included hydrocyanic acid and acetonitrile (345), in addition to those found in a steady state system. The reaction of hydrogen radicals with ethyleneimine results in the formation of hydrocyanic acid in addition to methane (350). Important processes in the photolysis of ethyleneimine are nitrene extrusion and homolysis of the $\text{N}-\text{H}$ bond, as suggested and simulated by *ab initio* SCF calculations (351). The occurrence of ethyleneimine as an intermediate in the photolytic formation of hydrocyanic acid from acetylene and ammonia in the atmosphere of the planet Jupiter has been postulated (352), but is disputed (353).

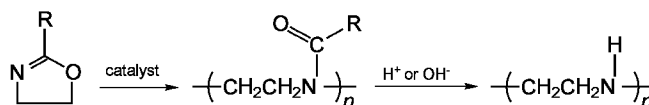
Polymerization. The polymerization of aziridines takes place in the presence of catalytic amounts of acid at elevated temperatures. The molecular weight can be controlled by the monomer–catalyst ratio, the addition of amines as stoppers, or the use of bifunctional initiators. In order to prevent a vigorous reaction, the heat liberated during the highly exothermic polymerization must be removed by various measures, ie, suitable dilution, controlled metering of the aziridine component, or external cooling after the reaction has started.

The polymerization of ethyleneimine (16,354–357) is started by a catalytically active reagent (H^+ or a Lewis acid), which converts the ethyleneimine into a highly electrophilic initiator molecule. The initiator then reacts with nitrogen nucleophiles, such as the ethyleneimine monomer and the subsequently formed oligomers, to produce a branched polymer, which contains primary, secondary, and tertiary nitrogen atoms in random ratios. Termination takes place by intramolecular macrocycle formation.



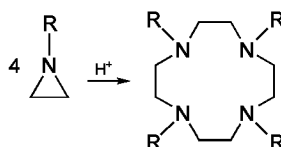
These branched polyethyleneimines are marketed, or have been marketed in the past, under the following trade names: Polymim (BASF Group, Germany and United States), Epomin (Nippon Shokubai, Japan), Corcat (Cordova Chemical Company, United States), and PEI (Dow Chemical Company, United States). In general, the reaction conditions employed for the polymerization of ethyleneimine do not have a significant effect on the degree of branching in the resulting polymers, so polyethyleneimines with extremely high degrees of branching are synthesized by other routes.

Linear polyethyleneimine results only in low yields from low temperature polymerization of ethyleneimine for very long reaction times. It can, however, be synthesized in a targeted manner by polymerization of 2-oxazolines with subsequent hydrolytic cleavage of the resulting polyamides (355,358).



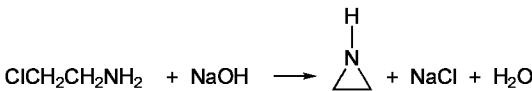
The polymerization of *N*-(2-tetrahydropyranyl)aziridine with subsequent hydrolysis of the resulting polymers has been described as an alternative route for the synthesis of linear polyethyleneimine (359). Linear polyethyleneimine, in contrast to branched polyethyleneimines, is only sparingly soluble in water at room temperature.

Polyethyleneimines with mainly tertiary and primary amino groups are synthesized by subjecting tris(2-aminoethyl)amine, as initiator core, to peraminoalkylation with methanesulfonylaziridine, and the methanesulfonyl protective groups are then removed by hydrolysis. Spherical starburst dendrimers, which contain tertiary amino groups on the inside and primary amino groups on the surface, are obtained by repeating this synthesis sequence several times (360). The reaction of 1-alkylaziridines, in particular 1-benzylaziridines, with catalytic amounts of acids in ethanol leads to the formation of cyclic oligomers, which are normally composed of four aziridine units (361–363).

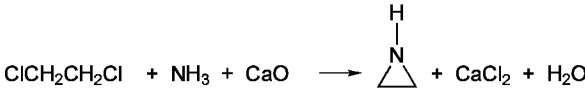


Preparation

Ethyleneimine was first prepared commercially in 1938 from 2-chloroethylamine and NaOH by a modified Gabriel synthesis (21).

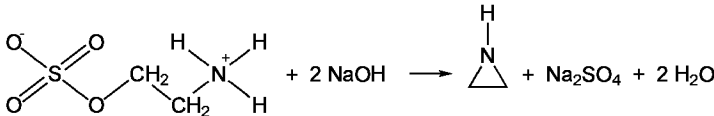


In the 1960s and 1970s ethyleneimine was produced by the dichloroethane–ammonia process by the Dow Chemical Co.

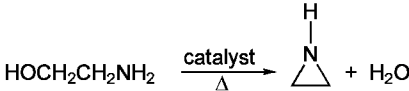


The problems inherent to these two processes are not only the production of corrosive salts, but also the possibility of product contamination by 2-chloroethylamine [689-98-5], as starting material or intermediate. This substance can initiate polymerization of ethyleneimine with the elimination of HCl.

The Wenker process (364), carried out by BASF and various other companies since the end of the 1960s, is a distinct improvement. In this process the hemisulfate of monoethanolamine, a nonvolatile, crystalline substance, is used in place of volatile 2-chloroethylamine for the alkaline cyclization. The reaction can be carried out under pressure (365).



A production plant for salt-free ethyleneimine synthesis by catalytic dehydration of monoethanolamine [141-45-5] in the gas phase has started operation at the Japanese company Nippon Shokubai (366).



Economic Aspects

Because of its toxicity, ethyleneimine monomer is not sold by the BASF group or Nippon Shokubai, currently the only large producers. Ethylenimine is used on-site for further reaction to produce polymers and intermediates. The BASF Corp. has started production of ethyleneimine in the United States in Freeport, Texas. The current world ethyleneimine production capacity is more than 12,000 t yr. In the United States alone, the consumption of polyethyleneimines approximately doubled between 1983 and 1990. One of the largest markets worldwide is the paper industry, which uses ethyleneimine polymers and their derivatives as process chemicals for paper making. Large amounts of ethyleneimine-based polymers are also used as oil field chemicals and flocculating (aggregating) agents. In addition, ethyleneimine derivatives can be used for a large number of special applications, such as enzyme immobilization, textile finishing, membranes, etc.

Analytical and Detection Methods, Storage

In addition to modern spectroscopic methods (¹H nmr spectroscopy, fir spectroscopy) and chromatographic methods (gc, hplc), HBr titration (29) is suitable for the quantitative analysis of ethyleneimine samples which contain relatively large amounts of ethyleneimine. In this titration, the ethyleneimine ring is opened with excess HBr in glacial acetic acid, and unconsumed HBr is back-titrated against silver nitrate.

Aziridine traces in aqueous solution can be determined by reaction with 4-(*p*-nitrobenzyl)pyridine [1083-48-3] and potassium carbonate [584-08-7]. Quantitative determination is carried out by photometric measurement of the absorption of the blue dye formed (367,368). Alkylating reagents interfere in the determination. Aziridine traces in the air can be detected discontinuously by absorption in Folin's reagent (1,2-naphthoquinone-4-sulfonate) [2066-93-5] (369,370) with subsequent chloroform extraction and hplc analysis of the red dye formed (371,372). The detection limit is ca 0.1 ppm. Nitrogen-specific thermal ionization detectors can be used for continuous monitoring of the ambient air.

Storage. Resistant materials for the storage and handling of ethyleneimine are low carbon steel, V2A and V4A chrome nickel steel, and (strengthened) glass and enamel. Copper, silver, and solders containing these metals must be avoided since they are attacked by ethyleneimine and can form potentially explosive compounds. The majority of plastics swell on contact with ethyleneimine and are unsuitable for its storage. Polytetrafluoroethylene can be used as sealing material for a limited period. When storing ethyleneimine, contact with acids, acid-producing substances, oxidizing agents, and sources of ignition must be avoided in order to avoid uncontrolled decomposition, which in the worst case can proceed with explosive force. Nitrogen is a suitable blanketing gas for use when storing ethyleneimine, and is used to prevent contact with the carbon dioxide in the air. Storage tanks should be fitted with a temperature monitoring device and a linked facility for NaOH metering.

Health and Safety Factors

Toxicology. Ethyleneimine is highly toxic on inhalation, on contact with the skin, and if swallowed (373). Inhalation of ethyleneimine vapors can cause damage to the mucous membranes of the respiratory tract and bronchi, with associated difficulty in breathing and irritation of the throat, and pulmonary edema. There can be a latency period of several hours between inhalation and onset of symptoms. Ethyleneimine vapors and liquid ethyleneimine can cause damage to the eyes, which can lead to blindness. Liquid ethyleneimine penetrates the skin very rapidly and produces severe burns and necroses (localized death of living tissue). In addition to the local effect, absorption of ethyleneimine can cause nausea and vomiting and states of agitation with profuse sweating, frequently only after a latency period of several hours. In the extreme case, ethyleneimine poisoning can be fatal.

The odor threshold for detection of ethyleneimine is 2 ppm. The maximum permissible concentration of ethyleneimine in the air at the place of work is 0.5 ppm (as specified in statutory regulations in the United States (374) and in Germany (375)). Animal experiments have shown ethyleneimine to be both carcinogenic (376) and mutagenic (377) (Table 2).

Table 2. Acute Toxicities for Ethyleneimine

Mode of administration	Species	LD ₅₀	Reference
inhalation (10 min),	mouse	4 mg L/air	378
injection,	rat	4 mg kg	379
ingestion,	rat	15 mg kg	380
absorption through the skin,	guinea pig	14 μL kg	381

Handling Precautions. It is essential that inhalation of vapors and contact with the skin is avoided when handling ethyleneimine. Suitable

personal protection includes a full protection suit, preferably made of butyl rubber, and a breathing and face mask (hood with independent air supply). Reactions with ethyleneimine carried out in the laboratory must be conducted as far as possible in closed apparatus in an efficient hood. Ethyleneimine and other low molecular weight aziridines are very highly flammable and can form explosive mixtures with air (Table 1). Possible sources of ignition (open flames, electric sparks, static charges, etc) must be removed when using ethyleneimine. Aziridines stored in the monomer form should be stabilized with solid NaOH to prevent spontaneous polymerization by traces of acid, eg, by carbon dioxide from the air. Ethyleneimine can be destroyed by slow introduction into sodium bisulfite solution.

Applications

Ethyleneimine and its derivatives have many industrial applications. Polymers based on ethyleneimine have a very high nitrogen:carbon ratio. This results in a high cation activity, which results in a high affinity for naturally occurring anions (29). However, the potential of ethyleneimine is based not only on autocondensation but also on the aminoethylation of functional groups. By grafting ethyleneimine onto oligomers or polymers, it is possible to adjust the cation activity of these compounds, and to tailor the compounds to suit the intended application (1,29).

The biological activity of ethyleneimine derivatives is utilized in both medicine and in crop protection (1). Derivatization of the aziridine ring results in a great variety of useful compounds. Derivatives of methylaziridine have a higher affinity for lipophilic substrates, and thus have widespread use as biologically active substances. The complexing properties of polyaziridines can be modified by the use of ethyleneimine derivatives such as *N*-(2-hydroxyethyl)aziridine as starting monomers (1).

Paper Industry. The principal use of ethyleneimine is as polymer in paper making. Polyethyleneimines (PEIs) promote the flocculation of wood and cellulose fibers and thus increase the retention of the fibers and fillers (382–384). The introduction of paper making in a neutral medium and the associated reduced use or abandonment of aluminum sulfate is increasing the demand for synthetic retention agents (385). Polyamidoamines are also suitable for this application, and the nitrogen content of these compounds can be increased by grafting with ethyleneimine (386–388). Because of their high positive charge density, PEIs fix negatively charged resin particles and interfering substances and thus facilitate problem-free paper production even in the case of closed water circulation (389,390), which is increasingly used. These fixing properties also can increase the quality of the paper. Ink-jet paper, which has to absorb ink easily, is produced with the addition of PEI (391). The fixing of sizing emulsions also serves the same purpose (392). In addition, both dry compaction (393) and wet compaction (394) are improved by the addition of polyethyleneimines.

Extraction and Complexing. The basic nitrogen atoms in PEI are able to absorb gases. This is utilized in, for example, cigarette filters, which can be impregnated with PEIs or derivatives (395), to remove aldehydes by chemical absorption (396). Acid gases can also be neutralized and absorbed by cross-linked PEIs (397) or by carriers impregnated with PEI (398). The complex-forming properties of PEI can also be used for the retention of metal ions (399–401) and for the catalysis of chemical reactions (402,403).

PEI derivatives have proven to be effective carriers of cations in liquid membrane systems (404). This technology led to the development of ion-exchange resins (405), which are also suitable for extracting uranium from seawater (406).

Polyelectrolytes based on ethyleneimine are also used to treat drinking water and process water, and as agents for preventing lime deposits (407) in water extraction. The binding power of PEI is utilized for the treatment of effluents (408). Biochemical reactions can be catalyzed by using the complex-forming properties of PEIs and their affinity for organic substrates (409).

Additives. Because of their versatility, imparted via chemical modification, the applications of ethyleneimine encompass the entire additive sector. The addition of PEI to PVC plastisols increases the adhesion of the coatings by selective adsorption at the substrate surface (410). PEI derivatives are also used as adhesion promoters in paper coating (411). The adducts formed from fatty alcohol epoxides and PEI are used as dispersants and emulsifiers (412). They are able to control the viscosity of dispersions, and thus facilitate transport in pipe systems (413). Fatty acid derivatives of PEI are even able to control the viscosity of pigment dispersions (414). The high nitrogen content of PEIs has a flame-retardant effect. This property is used, in combination with phosphorus compounds, for providing wood panels (415), cellulose (416), or polymer blends (417,418) with a flame-retardant finish.

Immobilization. The fixing property of PEIs has previously been discussed. Another application of this property is enzyme immobilization (419). Enzymes can be bound by reactive compounds, eg, isothiocyanate (420) to the PEI skeleton, or immobilized on solid supports, eg, cotton by adhesion with the aid of PEIs. In every case, fixing considerably simplifies the performance of enzyme-catalyzed reactions, thus facilitating preparative work. This technique has been applied to glutaraldehyde-sensitive enzymes (421), α -glucose transferase (422), and pectin lyase, pectin esterase, and endopolygalacturonase (423).

Substances other than enzymes can be immobilized. Examples include the fixing of heparin on polytetrafluoroethylene with the aid of PEI (424), the controlled release of pesticides which are bound to PEI (425), and the inhibition of herbicide suspensions by addition of PEI (426). The uptake of anionic dyes by fabric or paper is improved if the paper is first cationized with PEI (427). In addition, PEI is able to absorb odorizing substances such as fatty acids and aldehydes. Because of its high molecular weight, PEI can be used in cosmetics and body care products, as well as in industrial elimination of odors, such as the improvement of ambient air quality in sewage treatment plants (428).

Textile Finishing. Polyethyleneimine-*N*-methylolurea derivatives improve the crease and wear resistance of cotton (429,430). The adhesion between individual wool fibers is improved by pretreatment with amines, which leads to improved shrink resistance (431). An antimicrobial finish can be applied to cotton by using a combination of PEI and ureas to bind zinc pyrithione to the fabric (432). After wool has been provided with a flameproof finish using fluorozirconate or fluorotitanate, the wool can be neutralized with PEI (433). Conventional neutralizing agents cannot be used for this purpose since they impair the flameproof characteristics of the impregnated fabric.

Flocculation. The interaction of the cationic PEIs with anionic substrates leads to substrate flocculation. Applications of this property include the coagulation of latex (434), commercial application in effluent treatments (435–437), and stabilization of highly loaded coal–water mixtures in mining (438).

Material Protection. The graft copolymers of ethylene sulfide on polyethyleneimine can be used as an antifouling anticorrosion substrate for iron (439). PEIs or their derivatives are also used in electrolysis baths as brighteners in the electrochemical deposition of metals (440,441).

Medicine. There are many applications of PEI in the medical sector. Analytical methods, such as the quantitative determination of the surface charge of serum lipoproteins (442), are aided by use of PEI, and it is also used as a carrier in the development of polymer drugs (443,444). Platinum–polyethyleneimine complexes prevent the division of bacteria, and are being tested as carriers in the treatment of cancer and viruses (445–447). Encapsulated PEIs containing nucleic acid bases activate the neutrophils in human blood (448).

Aziridines occur naturally in the form of mitomycins (Table 3), which have antibiotic activity (1,449). Mytomycin C is used clinically as one of the most effective agents in the chemotherapy of cancer (450).

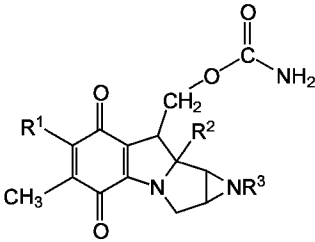


Table 3. Structures of Mytomycins

Name	CAS Registry Number	R ¹	R ²	R ³
------	---------------------	----------------	----------------	----------------

mitomycin A	[4055-40-7]	CH ₃ O	CH ₃ O	H
mitomycin B	[4055-39-4]	CH ₃ O	HO	CH ₃
mitomycin C	[50-07-7]	H ₂ N	CH ₃ O	H
porfiromycin	[801-52-5]	H ₂ N	CH ₃ O	CH ₃

Membranes and Osmosis. Membranes based on PEI can be used for the dehydration of organic solvents such as 2-propanol, methyl ethyl ketone, and toluene (451), and for concentrating seawater (452–454). On exposure to ultrasound waves, aqueous PEI salt solutions and brominated poly(2,6-dimethylphenylene oxide) form stable emulsions from which it is possible to cast membranes in which submicrometer capsules of the salt solution are embedded (455). The rate of release of the salt solution can be altered by surface–active substances. In membranes, PEI can act as a proton source in the generation of a photocurrent (456). The formation of a PEI coating on ion-exchange membranes modifies the transport properties and results in permanent selectivity of the membrane (457). The electrochemical testing of salts (458) is another possible application of PEI.

Miscellaneous Applications. PEIs and their derivatives are used as cementation auxiliaries in crude oil exploration (459), and for breaking crude oil emulsions (460) in crude oil extraction. Seed coatings of water-soluble copolymers containing polyethyleneimine have been developed (461). Polyethyleneimine derivatives have positive photoresist properties (462); amidated polyethyleneimines improve the flow properties of cement (463); and with few exceptions, *N*-acylaziridines act as chemical sterilizers for insects (464).

Derivatives. The most important data for 2-methylaziridine and 1-(2-hydroxyethyl)aziridine, which previously had some industrial significance, are given in Table I. Like ethyleneimine, these compounds are used in polymer form and as intermediates. The use of activated aziridines, eg, *N*-acylaziridines, for amino ethylation, under alkaline conditions, is of preparative interest (1).

BIBLIOGRAPHY

"Imines, Cyclic" in *ECT* 2nd ed., Vol. 11, pp. 526–548, by O. C. Dermer, Oklahoma State University, and A. W. Hart, The Dow Chemical Co.; in *ECT* 3rd ed., Vol. 13, pp. 142–166, by G. E. Ham, Dow Chemical U.S.A.

- O. C. Dermer and G. E. Ham, *Ethylenimine and Other Aziridines*, Academic Press, Inc., New York, 1969.
- H. Bestian, *Ann. Chem.* **566**, 210 (1950).
- H. Bestian, in E. Müller, ed., *Methoden der Organischen Chemie (Houben-Weyl)*, Georg Thieme Verlag, Stuttgart, 1958, p. 227.
- P. E. Fanta, in A. Weissberger, ed., *Heterocyclic Compounds with Three- and Four-Membered Rings*, Part I, Interscience Publishers, New York, 1964, p. 524.
- P. A. Gembitskii, N. M. Loim, and D. S. Zhuk, *Russ. Chem. Rev.* **35**, 105 (1966).
- S. Hirai and W. Nagata, *Kagaku No Ryoiki Zokan* **87**, 213 (1969).
- F. N. Gladysheva, A. P. Sineokov, and V. S. Etlis, *Russ. Chem. Rev.* **39**, 118 (1970).
- H. W. Heine, *Mech. Mol. Migr.* **3**, 145 (1971).
- J. W. Lown, *Rec. Chem. Progr.* **32**, 51 (1971).
- J. W. Lown and K. Matsumoto, *Yuki Gosei Kagaku Kyokai Shi* **29**, 760 (1971).
- G. Maerker, in F. D. Gunstone, ed., *Topics in Lipid Chemistry*, Vol. 2, Logos Press Limited, London, 1971, p. 159.
- G. E. Ham, in K. C. Frisch, ed., *High Polymers*, Vol. 26, John Wiley & Sons, Inc., New York, 1972, p. 313.
- G. J. Matthews and A. Hassner, in J. Fried and J. A. Edwards, eds., *Organic Reactions in Steroid Chemistry*, Vol. 2, 1972, p. 1.
- O. Cervinka, *Method Chim.* **6**, 603 (1975).
- G. E. Ham, in N. M. Bikales, ed., *Encyclopedia of Polymer Science and Technology*, Suppl. Vol. 1, John Wiley & Sons, Inc., New York, 1976, p. 25.
- G. E. Ham, in E. J. Goethals, ed., *Polymeric Amines and Ammonium Salts*, Pergamon Press, New York, 1980, p. 1.
- M. Kodo, *Kagaku Keizai*, 41 (1982).
- J. A. Deyrup, in A. Weissberger and E. C. Taylor, eds., *The Chemistry of Heterocyclic Compounds*, Vol. 42, Part 1, John Wiley & Sons, Inc., New York, 1983, p. 1.
- A. Padwa and A. D. Woolhouse, in A. R. Katritzky and C. W. Rees, eds., *Comprehensive Heterocyclic Chemistry*, Vol. 7, Pergamon Press, Oxford, U.K., 1984, p. 47.
- S. Gabriel, *Ber.* **21**, 1049 (1988).
- Ibid.*, p. 2664.
- C. C. Howard and W. Marckwald, *Ber.* **32**, 2036 (1899).
- W. Marckwald, *Ber.* **33**, 764 (1900).
- P. F. Pascoe and W. A. Sherbrock-Cox, *J. Appl. Chem.* **13**, 564 (1963).
- Sorbe, *Sicherheitstechnische Kenndaten Chemischer Stoffe*, Ecomed Verlagsgesellschaft GmbH, Landsberg Lech, 1983, p. 11.
- Internal technical research report, *Ethylenimin*, BASF AG, Ludwigshafen, Germany, 1970.
- E. W. Flick, ed., *Industrial Solvents Handbook*, 3rd ed., Noyes Data Corp., Park Ridge, N.J., 1985, p. 569.
- Ref. 25, p. 9.
- Technical data, BASF AG, Ludwigshafen, Germany, 1964.
- R. A. Nelson and R. S. Jessup, *J. Res. Natl. Bur. Standards* **48**, 206 (1952).
- Vapour Pressures and Critical Points of Liquids XI: Heterocyclic Nitrogen Compounds*, Engineering Sciences Data Unit 77019, London, 1977.
- D. R. Stull, E. F. Westrum, Jr., and G. C. Sinke, *The Chemical Thermodynamics of Organic Compounds*, John Wiley & Sons, Inc., New York, 1969, pp. 468–469, and references therein.
- J. Hinze and R. F. Curl, Jr., *J. Am. Chem. Soc.* **86**, 5068 (1964).
- U.S. Pat. 2,729,679** (Nov. 19, 1951), J. L. Anderson (to E. I. du Pont de Nemours & Co., Inc.).
- K. Hashigushi and co-workers, *J. Mol. Spectroscopy* **105**, 81 (1984).
- J. L. Ripoll, H. Lebrun, and A. Thuillier, *Tetrahedron* **36**, 2497 (1980).
- Y. Hamada and co-workers, *J. Mol. Spectroscopy* **105**, 93 (1984).
- R. I. Wagner and co-workers, *J. Am. Chem. Soc.* **89**, 1102 (1967).
- G. D. Jones and co-workers, *J. Org. Chem.* **9**, 125 (1944).
- L. Siggel and S. Brode, technical data, BASF AG, Ludwigshafen, Germany, 1992.
- R. Hoffmann, *Tetrahedron Lett.*, 2907 (1970).
- R. Hoffmann and W. D. Stohrer, *J. Am. Chem. Soc.* **93**, 6941 (1971).
- R. Gleiter, *Topics Curr. Chem.* **86**, 197 (1979).
- W. S. Wilcox and co-workers, *J. Chem. Phys.* **21**, 563 (1953).
- T. E. Turner and co-workers, *J. Chem. Phys.* **21**, 564 (1953).
- T. G. Tuner, V. C. Flora, and W. M. Kendrick, *J. Chem. Phys.* **23**, 1966 (1955).
- R. D. Johnson, R. J. Myers, and W. D. Gwinn, *J. Phys. Chem.* **21**, 1425 (1953).
- M. Igarashi, *Bull. Chem. Soc. Jpn.* **34**, 369 (1961).
- H. B. Basch and co-workers, *J. Chem. Phys.* **51**, 52 (1969).
- P. Rademacher and co-workers, *J. Mol. Struct.* **197**, 291 (1989).

51. R. Ellerbrock and P. Rademacher, *J. Mol. Struct.* **265**, 93 (1992).

52. C. Liang and L. C. Allen, *J. Am. Chem. Soc.* **113**, 1878 (1991) and references therein.

53. G. J. Buist and H. J. Lucas, *J. Am. Chem. Soc.* **79**, 6157 (1957).

54. D. R. Crist and N. J. Leonard, *Angew. Chem.* **81**, 953 (1969) and references therein.

55. U. Harder, E. Pfeil, and K. F. Zenner, *Chem. Ber.* **97**, 510 (1964).

56. H. Stamm, *Angew. Chem.* **74**, 694 (1962).

57. H. Stamm, *Angew. Chem.* **77**, 738 (1965).

58. H. Stamm, *Chem. Ber.* **99**, 2556 (1966).

59. H. Stamm, *Arch. Pharm.* **299**, 965 (1966).

60. G. I. Bras, *Dokl. Akad. Nauk SSSR* **87**, 589 (1952).

61. G. I. Bras, *Dokl. Akad. Nauk SSSR* **87**, 747 (1952).

62. N. Milstein, *J. Heterocycl. Chem.* **5**, 339 (1968).

63. H. Stamm and co-workers, *J. Org. Chem.* **54**, 193 (1989).

64. R. Bucourt and M. Vignau, *Bull. Soc. Chim. Fr.*, 1190 (1961).

65. E. Pfeil and U. Harder, *Angew. Chem.* **79**, 188 (1967).

66. J. Legters, L. Thijs, and B. Zwanenburg, *Recl. Trav. Chim. Pays-Bas* **111**, 16 (1992).

67. R. Fikentscher, R. Brückmann, and R. Betz, *Ann. Chem.*, 113 (1990).

68. E. D. Schilling and F. M. Strong, *J. Org. Chem.* **22**, 349 (1957).

69. S. Matsubara, T. Kodama, and K. Utimoto, *Tetrahedron Lett.* **31**, 6379 (1990).

70. K. Utimoto and co-workers, *Yuki Gosei Kagaku Kenkyusho Koenshu* **5**, 101 (1991).

71. L. B. Clapp, *J. Am. Chem. Soc.* **70**, 184 (1948); **Brit. Pat. 1,363,299** (Nov. 26, 1970), F. Mikovsky and R. Fikentscher (to BASF AG); **Brit. Pat. 1,383,127** (Jan. 11, 1971), F. Mikovsky and R. Fikentscher (to BASF AG).

72. A. R. Dalin and co-workers, *Zh. Obshch. Khim.* **58**, 2098 (1988).

73. **Eur. Pat. 422,534 A2** (Jan. 19, 1994), A. Hohmann, W. Reuther, and W. Bochnitschek (to BASF AG).

74. G. I. Bras and V. A. Skorodumov, *Compt. Rend. Acad. Sci. USSR* **55**, 315 (1947).

75. G. I. Bras and V. A. Skorodumov, *Dokl. Akad. Nauk SSSR* **59**, 489 (1948).

76. G. H. Coleman and J. E. Callen, *J. Am. Chem. Soc.* **68**, 2006 (1946).

77. **Eur. Pat. 370,730 A2** (Nov. 22, 1988), Y. Hino and co-workers, (to Nippon Shokubai Kagaku Kogyo Co., Ltd.).

78. **Ger. Pat. 1,108,223** (Oct. 9, 1959), K. Eiter and E. Truscheit (to Bayer AG).

79. R. D. Guthrie and co-workers, *Proc. Chem. Soc.*, 84 (1963).

80. E. Bertele and co-workers, *Angew. Chem.* **76**, 393 (1964).

81. G. Swift and D. Swern, *J. Org. Chem.* **32**, 511 (1967).

82. J. Cleophas, S. D. Gero, and J. Hildesheim, *Chem. Commun.*, 94 (1968).

83. Y. Hata and M. Watanabe, *Tetrahedron* **30**, 3569 (1974).

84. **Ger. Pat. 1,198,471** (Aug. 30, 1961), A. Tartter, G. Lützel, and W. Rohland (to BASF AG).

85. R. T. Markiw, *J. Org. Chem.* **37**, 2165 (1972).

86. M. E. Freed and A. R. Day, *J. Org. Chem.* **25**, 2108 (1960).

87. H. B. Sullivan and A. R. Day, *J. Org. Chem.* **29**, 326 (1964).

88. C. R. Rasmussen and D. L. Shaw, *J. Org. Chem.* **39**, 1560 (1974).

89. **U.S. Pat. 3,502,654** (Apr. 12, 1967), J. D. Young (to Dow Chemical Co.).

90. C. R. Dick, *J. Org. Chem.* **32**, 72 (1967).

91. W. Reuther, technical data, BASF AG, Ludwigshafen, Germany, 1989.

92. K. Issleib and D. Haferburg, *Z. Naturforsch.* **B 20**, 916 (1965).

93. U. Harder, E. Pfeil, and K. F. Zenner, *Chem. Ber.* **97**, 510 (1964).

94. **U.S. Pat. 2,703,816** (Sept. 26, 1951), W. S. Emerson and B. S. Wildi (to Monsanto Chemical Co.).

95. L. B. Clapp, *J. Am. Chem. Soc.* **73**, 2584 (1951).

96. R. Ghirardelli and H. J. Lucas, *J. Am. Chem. Soc.* **77**, 106 (1955).

97. A. Hassner and C. Heathcock, *J. Org. Chem.* **30**, 1748 (1965).

98. **U.S. Pat. 3,336,358** (July 23, 1964), R. T. McFadden (to Dow Chemical Co.).

99. **U.S. Pat. 3,338,954** (Oct. 8, 1964), R. T. McFadden and C. R. Dick (to Dow Chemical Co.).

100. L. B. Clapp and co-workers, *J. Am. Chem. Soc.* **77**, 5116 (1955).

101. **Ger. Pat. 1,081,473** (Mar. 18, 1954), H. G. Trieschmann, L. Reuther, and W. Arend.

102. **Brit. Pat. 784,058** (Mar. 18, 1954), W. Arendt, L. Reuter, and H. Trieschmann (to BASF AG).

103. **Brit. Pat. 784 059** (Mar. 18, 1954), W. Arendt, L. Reuter, and H. Trieschmann (to BASF AG).

104. P. Loewigkeit and co-workers, *J. Org. Chem.* **33**, 3344 (1968).

105. P. Golding and co-workers, *Tetrahedron Lett.* **32**, 4985 (1991).

106. P. Golding and co-workers, *Tetrahedron Lett.* **29**, 2735 (1988).

107. D. M. Brown and G. O. Osborne, *J. Chem. Soc.*, 2590 (1957).

108. H. N. Christensen, *J. Biol. Chem.* **135**, 399 (1940).

109. Y. S. Tsizin and N. A. Preobrazhenskii, *J. Gen. Chem. USSR* **33**, 2800 (1963).

110. S. Z. Ivin and co-workers, *J. Gen. Chem. USSR* **37**, 1556 (1967).

111. J. E. Eadley, *J. Am. Chem. Soc.* **80**, 3458 (1958).

112. S. Gabriel and G. Eschenbach, *Ber.* **30**, 2494 (1897).

113. **Eur. Pat. 320,727 A2** (Dec. 12, 1987), A. Hohmann and co-workers (to BASF AG).

114. R. G. Kostyanovskii, *Dokl. Akad. Nauk SSSR* **135**, 853 (1960).

115. J. L. Larice, J. Roggero, and J. Metzger, *Bull. Soc. Chim. Fr.*, 3637 (1967).

116. R. A. Wohl and D. F. Headley, *J. Org. Chem.* **37**, 4401 (1972).

117. G. Tsatsas, C. Sandris, and D. Kontonassios, *Bull. Soc. Chim. Fr.*, 3100 (1964).

118. H. Rübner, *Arzneim. Forsch.* **15**, 796 (1965).

119. G. Meguerian and L. B. Clapp, *J. Am. Chem. Soc.* **73**, 2121 (1951).

120. H. Lehr, S. Kadan, and M. W. Goldberg, *J. Med. Chem.* **6**, 136 (1963).

121. F. Asinger, F. J. Schmitz, and S. Reichel, *Ann. Chem.* **652**, 50 (1962).

122. F. Asinger, H. Diem, and W. Schäfer, *Monatsh. Chem.* **95**, 1335 (1964).

123. F. Asinger and co-workers, *Monatsh. Chem.* **101**, 1281 (1970).

124. *Ibid.*, p. 1295.

125. **U.S. Pat. 3,793,370** (Apr. 21, 1970), F. Mikovsky and R. Fikentscher (to BASF AG).

126. S. L. Holton and co-workers, *Can. J. Chem.* **42**, 2147 (1964).

127. L. D. Spicer and co-workers, *J. Org. Chem.* **33**, 1350 (1968).

128. **Ger. Pat. 893,795** (Apr. 26, 1951), R. Kuhn and G. Quadbeck (to BASF AG).
129. S. Akerfeldt, *Acta Chem. Scand.* **17**, 329 (1963).
130. P. Rumpf, *Compt. Rend.* **204**, 592 (1937).
131. F. C. Schaefer, J. T. Goegehegan, and D. W. Kaiser, *J. Am. Chem. Soc.* **77**, 5918 (1955).
132. E. Allen and W. Seaman, *Anal. Chem.* **27**, 540 (1955).
133. H. Distler, *Angew. Chem.* **79**, 520 (1967).
134. J. E. Easley and co-workers, *J. Am. Chem. Soc.* **80**, 3458 (1958).
135. S. Akerfeldt and L. Fagerlind, *J. Med. Chem.* **10**, 115 (1966).
136. F. Asinger, H. Berding, and H. Offermanns, *Monatsch. Chem.* **99**, 2084 (1968).
137. D. L. Klayman, *J. Org. Chem.* **30**, 2454 (1965).
138. A. Barama, R. Condom, and R. Guedj, *J. Fluorine Chem.* **16**, 183 (1980).
139. T. N. Wade and R. Guedj, *Tetrahedron Lett.*, 3247 (1978).
140. T. N. Wade, F. Gaymard, and R. Guedj, *Tetrahedron Lett.*, 2681 (1979).
141. **Rus. Pat. 244,339** (Feb. 27, 1967), M. Woronkow, L. Fedotowa, and D. Rinkis (to Institut für organische Synthese der Akademie der Wissenschaften der Lehsichen sozialistischen sowjet Republik).
142. R. C. Elderfield and H. A. Hageman, *J. Org. Chem.* **14**, 605 (1949).
143. J. L. Atwood and G. D. Stucky, *J. Am. Chem. Soc.* **92**, 285 (1970).
144. A. Storr and B. S. Thomas, *J. Chem. Soc. (A)*, 3850 (1971).
145. A. Storr, B. S. Thomas, and A. D. Penland, *J. Chem. Soc. Dalton Trans.*, 326 (1972).
146. J. Möller, K. Margiolis, and K. Dehnicke, *J. Organomet. Chem.* **46**, 219 (1972).
147. H. C. Brown and M. Gerstein, *J. Am. Chem. Soc.* **72**, 2926 (1950).
148. B. P. Robinson and K. A. H. Adams, *Tetrahedron Lett.*, 6169 (1968).
149. S. Akerfeldt, K. Wahlberg, and M. Hellström, *Acta Chem. Scand.* **23**, 115 (1969).
150. H. Ringertz, *Acta Chem. Scand.* **23**, 137 (1969).
151. R. Ehdich and co-workers, *Inorg. Chem.* **3**, 628 (1944).
152. **U.S. Pat. 3,393,184** (Dec. 4, 1964), J. A. Hoffman (to American Cyanamid Co.).
153. M. G. Voronkov and L. A. Fedotova, *Khim. Geterotsikl. Soedin.* **2**, 408 (1966).
154. **U.S. Pat. 3,197,463** (Mar. 10, 1961), G. C. Tesoro, D. Ferry, and S. B. Sello (to J. P. Stevens and Co.).
155. M. T. Maurette and co-workers, *C.R. Acad. Sci. Paris, Ser. C* **282**, 599 (1976).
156. R. G. Kostyanovskii and A. K. Prokof'ev, *Izv. Akad. Nauk SSSR, Ser. Khim.*, 159 (1965).
157. W. Funke, *Ann. Chem.* **725**, 15 (1969).
158. R. G. Kostyanovskii, Y. I. El'natanov, and K. Khafizov, *Izv. Akad. Nauk SSSR, Ser. Khim.*, 1815 (1970).
159. T. Kametani and co-workers, *Israel J. Chem.* **27**, 57 (1986).
160. S. S. Skorokhodov and co-workers, *Zh. Obshch. Khim.* **31**, 3382 (1961).
161. W. D. Stephens and co-workers, *J. Chem. Eng. Data* **8**, 625 (1963).
162. C. W. Woods, A. B. Borkovec, and F. M. Hart, *J. Med. Chem.* **7**, 371 (1964).
163. Y. Iwakura and A. Nabeya, *J. Org. Chem.* **25**, 1118 (1960).
164. H. W. Heine, *J. Am. Chem. Soc.* **85**, 2743 (1963).
165. H. W. Heine, D. C. King, and L. A. Portland, *J. Org. Chem.* **31**, 2662 (1966).
166. P. Reynaud, R. C. Moreau, and P. Fodor, *C.R. Acad. Sci. Paris, Ser. C* **266**, 632 (1968).
167. T. Nishiguchi and co-workers, *J. Am. Chem. Soc.* **91**, 5835 (1969).
168. T. A. Foglia, L. M. Gregory, and G. Maerker, *J. Org. Chem.* **35**, 3779 (1970).
169. J. F. Bieron and F. J. Dinan, in J. Zabicky, ed, *The Chemistry of Amides*, Wiley-Interscience, New York, 1970, p. 246.
170. F. A. Vingiello, M. P. Rorer, and M. A. Ogliaruso, *J. Chem. Soc., Chem. Commun.*, 329 (1971).
171. S. P. McManus, R. A. Hearn, and C. U. Pittman, Jr., *J. Org. Chem.* **41**, 1895 (1976).
172. V. P. Wystrach, D. W. Kaiser, and F. C. Schaefer, *J. Am. Chem. Soc.* **77**, 5915 (1955).
173. C. K. Johnson, *J. Org. Chem.* **32**, 1508 (1967).
174. D. A. Tomalia, N. D. Ojha, and B. P. Thill, *J. Org. Chem.* **34**, 1400 (1969).
175. D. A. Tomalia, *J. Heterocycl. Chem.* **3**, 384 (1966).
176. D. A. Tomalia, *J. Heterocycl. Chem.* **4**, 419 (1967).
177. D. A. Tomalia, *Tetrahedron Lett.*, 2559 (1967).
178. D. A. Tomalia, T. J. Giacobbe, and W. A. Sprenger, *J. Org. Chem.* **36**, 2142 (1971).
179. E. V. Dehmloew and H. J. Westendorf, *Z. Naturforsch.* **B 25**, 1191 (1970).
180. A. Funke and G. Benoit, *Bull. Chem. Soc. Chim. Fr.*, 1021 (1953).
181. **Ger. Pat. 1,004,614** (June 26, 1954), K. Vierling, H. Oettel, and G. Wilhelm (to BASF AG).
182. **Ger. Pat. 1,031,307** (Oct. 24, 1956), H. Metzger, K. Vierling, and G. Wilhelm (to BASF AG).
183. **Ger. Pat. 1,054,998** (Oct. 9, 1957), O. V. Schickh and co-workers, (to BASF AG).
184. **Ger. Pat. 1,064,500** (Sept. 5, 1956), H. Metzger and co-workers, (to BASF AG).
185. **Ger. Pat. 1,115,256** (Apr. 26, 1957), K. Oemmier and co-workers, (to BASF AG).
186. E. J. Reist, I. G. Junga, and B. R. Baker, *J. Org. Chem.* **25**, 1673 (1960).
187. E. J. Reist and co-workers, *J. Org. Chem.* **26**, 2139 (1961).
188. F. Fischer and H. Rünsch, *Chem. Ber.* **94**, 901 (1961).
189. P. W. Feit, *Acta Chem. Scand.* **16**, 522 (1962).
190. H. A. Staab and W. Rohr, *Chem. Ber.* **95**, 1298 (1962).
191. **Brit. Pat. 1,138,738** (Dec. 1, 1966), D. D. Reynolds (to Eastman Kodak Co.).
192. **Ger. Pat. 1,029,000** (Oct. 24, 1956), H. Metzger, H. Oettel, and G. Wilhelm (to BASF AG).
193. **U.S. Pat. 3,520,875** (Aug. 25, 1966), T. L. Ashby and R. E. Lane, Jr. (to Dow Chemical Co.).
194. **U.S. Pat. 2,654,737** (Sept. 26, 1949), H. Bestian (to Hoechst AG).
195. **Brit. Pat. 692,368** (Sept. 26, 1949), M. W. Edenbach and A. Sieglitz (to Hoechst AG).
196. J. M. Stewart, *J. Am. Chem. Soc.* **76**, 3228 (1954).
197. D. Rosenthal and co-workers, *J. Org. Chem.* **30**, 3689 (1965).
198. Y. Oshiro, T. Agawa, and S. Komori, *Osaka Daigaku (Technol. Rep. Osaka Univ.)*, 541 (1967).
199. Y. I. El'natanov and R. G. Kostyanovskii, *Izv. Akad. Nauk SSSR, Ser. Khim.*, 1661 (1988).
200. **Ger. Pat. 1,111,637** (Dec. 14, 1957), F. Reicheneder, K. Dury, and H. J. Sasse (to BASF AG).
201. J. E. Dolfini, *J. Org. Chem.* **30**, 1298 (1965).
202. E. I. Grinblat and I. Y. Postovskii, *Dokl. Akad. Nauk SSSR* **133**, 873 (1960).
203. I. A. Postovskii, E. I. Grinblat, and L. F. Trefilova, *Zh. Obshch. Khim.* **31**, 363 (1961).
204. K. C. Tsou, K. Hoegerle, and H. C. F. Su, *J. Med. Chem.* **6**, 435 (1963).

205. Y. I. El'natanov and R. G. Kostyanovskii, *Izv. Akad. Nauk SSSR*, 302 (1988).

206. **Ger. Pat. 1,010,063** (May 22, 1953), W. Riemenschneider (to Hoechst AG).

207. W. E. Truce and D. G. Brady, *J. Org. Chem.* **31**, 3543 (1966).

208. W. E. Truce and L. D. Markley, *J. Org. Chem.* **35**, 3275 (1970).

209. W. E. Truce and D. W. Onken, *J. Org. Chem.* **40**, 3200 (1975).

210. **U.S. Pat. 2,770,617** (Apr. 30, 1954), A. Marxer (to Ciba Pharmaceutical Products, Inc.).

211. T. S. Lin and co-workers, *J. Med. Chem.* **32**, 1467 (1989).

212. T. S. Lin and co-workers, *J. Med. Chem.* **34**, 1634 (1991).

213. J. E. Dolfini and D. M. Dolfini, *Tetrahedron Lett.*, 2053 (1965).

214. M. Fremery, *Angew. Chem.* **78**, 1024 (1966).

215. **U.S. Pat. 3,342,833** (Aug. 7, 1964), M. I. Fremery (to Shell Oil Co.).

216. **U.S. Pat. 3,259,632** (Aug. 7, 1964), M. I. Fremery (to Shell Oil Co.).

217. **U.S. Pat. 3,236,835** (Nov. 24, 1961), W. J. Rabourn and L. Jackson (to Dow Chemical Co.).

218. A. Dornow and W. Schacht, *Ber.* **82**, 464 (1949).

219. **Ger. Pat. 935,545** (Jan. 3, 1954), K. Vierling (to BASF AG).

220. W. J. Rabourn and W. L. Howard, *J. Org. Chem.* **27**, 1039 (1962).

221. R. G. Kostyanovskii and V. F. Bystrov, *Izv. Akad. Nauk SSSR* **8**, 1404 (1962).

222. R. G. Kostyanovskii and V. F. Bystrov, *Doklady Akad. Nauk SSSR* **148**, 97 (1963).

223. P. G. Kostyanovskii and O. A. Pan'shin, *Izv. Akad. Nauk SSSR*, 182 (1963).

224. S. C. Kuo and W. H. Daly, *J. Org. Chem.* **35**, 1861 (1970).

225. **Ger. Pat. 1,445,525** (Mar. 13, 1964), B. Sander and F. Becke (to BASF AG).

226. J. B. Doughty, C. L. Lazzell, and A. R. Collett, *J. Am. Chem. Soc.* **72**, 2866 (1950).

227. H. Fillion and co-workers, *Chem. Pharm. Bull.* **39**, 493 (1991).

228. Y. Ohshiro, K. Yamamoto, and S. Komori, *Yuki Gosei Kagaku Kyokai Shi* **24**, 945 (1966).

229. E. Pfeil and U. Harder, *Angew. Chem.* **77**, 505 (1965).

230. F. Lautenschlaeger, *J. Heterocycl. Chem.* **7**, 1283 (1970).

231. P. Rajagopalan and C. N. Talaty, *J. Am. Chem. Soc.* **88**, 5048 (1966).

232. T. Sasaki and T. Yoshioka, *Bull. Chem. Soc. Jpn.* **42**, 556 (1969).

233. T. Sasaki, T. Yoshioka, and Y. Suzuki, *Bull. Chem. Soc. Jpn.* **44**, 185 (1971).

234. R. Schwyzler, *Helv. Chim. Acta* **XXXV**, 1903 (1952).

235. **Ger. Pat. 1,087,131** (Oct. 11, 1958), W. Gauß and S. Petersen (to Bayer AG).

236. D. Bormann, *Angew. Chem. Int. Ed.* **12**, 768 (1973).

237. T. Kato, Y. Yamamoto, and M. Sato, *Yagugaku Zasshi* **91**, 384 (1971).

238. **Ger. Pat. 1,941,175** (Aug. 13, 1969), R. Fikentscher and F. Miksovsky (to BASF AG).

239. R. Kostyanovskii and co-workers, *Izv. Akad. Nauk SSSR*, 2576 (1988).

240. **U.S. Pat. 2,339,046** (Apr. 1, 1940), H. Bestian (to Hoechst AG).

241. L. B. Dashkevich and V. G. Beilin, *Zh. Obshch. Khim.* **32**, 2391 (1962).

242. N. Murai and co-workers, *J. Org. Chem.* **42**, 847 (1977).

243. B. Adcock and A. Lawson, *J. Chem. Soc.* **474** (1965).

244. A. B. Borkovec and C. W. Woods, *J. Med. Chem.* **8**, 545 (1965).

245. H. Najer and co-workers, *Compt. Rend.* **253**, 2369 (1961).

246. H. Najer and co-workers, *Bull. Soc. Chim. Fr.*, 323 (1963).

247. E. Pfeil and K. Milzner, *Angew. Chem.* **78**, 677 (1966).

248. A. S. Deutsch and P. E. Fanta, *J. Org. Chem.* **21**, 892 (1956).

249. J. C. Howard and G. Klein, *J. Org. Chem.* **27**, 3701 (1962).

250. Y. Iwakura and A. Nabeya, *Nippon Kagaku Zasshi* **77**, 773 (1956).

251. K. Soga and co-workers, *J. Chem. Soc., Chem. Commun.*, 617 (1976).

252. M. R. Banks and co-workers, *J. Chem. Soc. Perkin. Trans.*, 961 (1991).

253. H. Matsuda, A. Ninagawa, and H. Hasegawa, *Bull. Chem. Soc. Jpn.* **58**, 2717 (1985).

254. A. Seher, *Annalen* **575**, 153 (1952).

255. H. M. Rauen and H. Schriewer, *Arzneim. Forsch.* **21**, 518 (1971).

256. **Rus. Pat. 2,235,876 A2** (Sep. 18, 1990), M. Kitano and M. Yagisawa (to Nippon Shokubai Kagaku Kogyo Co., Ltd.).

257. V. I. Markov and D. A. Danileiko, *Zh. Org. Khim.* **10**, 1262 (1974).

258. L. B. Clapp and J. W. Watjen, *J. Am. Chem. Soc.* **75**, 1490 (1953).

259. C. S. Dewey and R. A. Bafford, *J. Org. Chem.* **30**, 491 (1965).

260. T. A. Foglia and co-workers, *J. Org. Chem.* **36**, 1068 (1971).

261. F. A. Devillanova and G. Verani, *Spectrochim. Acta* **36A**, 371 (1980).

262. S. F. Thames, J. E. McClesky, and P. L. Kelly, *J. Heterocycl. Chem.* **5**, 749 (1968).

263. **Ger. Pat. 834,990** (Oct. 2, 1948), J. Heyna and A. Bauer (to Hoechst AG).

264. **DDR 20,378** (Nov. 14, 1958), E. J. Birr and W. Walther.

265. N. S. Nametkin, V. N. Perchenko, and L. G. Batalova, *Dokl. Akad. Nauk SSSR* **158**, 953 (1964).

266. N. S. Nametkin, V. N. Perchenko, and G. K. Kadorkina, *Dokl. Akad. Nauk SSSR* **187**, 522 (1969).

267. N. S. Nametkin and co-workers, *Dokl. Akad. Nauk SSSR* **219**, 860 (1974).

268. N. S. Nametkin, V. N. Perchenko, and L. G. Batalova, *Dokl. Akad. Nauk SSSR* **160**, 167 (1965).

269. N. S. Nametkin and co-workers, *Khim. Geterotsikl. Soedin.* **3**, 75 (1967).

270. O. Scherer and M. Schmidt, *Chem. Ber.* **98**, 2243 (1965).

271. R. Riviere, M. Riviere-Baudet, and J. Satge, *J. Organomet. Chem.* **96**, C7 (1975).

272. J. Barrau and co-workers, *J. Organomet. Chem.* **243**, 281 (1983).

273. M. Riviere-Baudet and J. Satge, *Rec. Trav. Chem. Pays-Bas* **89**, 1281 (1970).

274. R. G. Kostyanovskii and A. K. Prokof'ev, *Izv. Akad. Nauk SSSR, Ser. Khim.*, 469 (1967).

275. M. E. Bishop and J. J. Zuckerman, *Inorg. Chem.* **16**, 1749 (1977).

276. W. Rundel and E. Möller, *Ber.* **96**, 2528 (1963).

277. R. D. Clark and G. K. Helmkamp, *J. Org. Chem.* **29**, 1316 (1964).

278. H. W. Heine and D. A. Tomalia, *J. Am. Chem. Soc.* **84**, 993 (1962).

279. H. W. Heine, *Angew. Chem.* **74**, 772 (1962).

280. L. A. Fedotova and M. G. Voronkov, *Khim. Geterotsk. Soed.* **3**, 570 (1969).

281. J. G. Pomonis, D. T. North, and R. G. Zaylskie, *J. Med. Chem.* **13**, 989 (1970).

282. **U.S. Pat. 2,858,306** (May 8, 1957) R. F. Ratz and C. J. Grundmann (to Olin Mathieson Chemical Corp.).

283. E. E. Nifant'ev, A. I. Zavalisina, and M. R. Ter-Ovanesjan, *Z. Obsc. Chim.* **39**, 28 (1969).
284. E. E. Nifant'ev, A. I. Gavrilova, and N. K. Bliznjuk, *Z. Obsc. Chim.* **36**, 363 (1966).
285. V. S. Abramov and N. A. Il'ina, *Z. Obsc. Chim.* **39**, 177 (1969).
286. R. Appel and co-workers, *Chem. Ber.* **103**, 3631 (1970).
287. G. Kamai and Z. L. Khisamova, *Dokl. Akad. Nauk SSSR* **156**, 485 (1964).
288. M. G. Voronkov, L. A. Fedotova, and D. O. Rinkis, *Khim. Geterotsikl. Soedin.*, 488 (1965).
289. F. Feher and B. Degen, *Angew. Chem.* **79**, 690 (1967).
290. F. Graefe and R. E. Meyer, *J. Am. Chem. Soc.* **80**, 3939 (1958).
291. J. M. Kanabus-Kaminska and J. Lange, *Rocz. Chem.* **51**, 1253 (1977).
292. W. Hennrichs and J. Jander, *Z. Anorg. Allg. Chem.* **547**, 233 (1987).
293. R. G. Jones and co-workers, *J. Am. Chem. Soc.* **78**, 6027 (1956).
294. H. Bärger and U. Dammgen, *Z. Anorg. Allg. Chem.* **394**, 209 (1972).
295. H. Bärger and U. Dammgen, *Z. Anorg. Allg. Chem.* **407**, 201 (1974).
296. A. D. Jenkins, M. F. Lappert, and R. C. Srivastava, *J. Organomet. Chem. (Lausanne)* **23**, 165 (1970).
297. T. B. Jackson and J. O. Edwards, *Inorg. Chem.* **1**, 398 (1962).
298. H. P. Fritz and G. Hierl, *Z. Anorg. Chem.* **371**, 229 (1969).
299. R. Hüfer, W. Beck, and A. Engelmann, *Chem. Ber.* **106**, 2590 (1973).
300. T. B. Jackson and J. O. Edwards, *J. Am. Chem. Soc.* **83**, 355 (1961).
301. I. Morishima and co-workers, *J. Am. Chem. Soc.* **102**, 2429 (1980).
302. W. Hieber and R. Wiesboeck, *Chem. Ber.* **91**, 1146 (1958).
303. R. W. Kiser and T. W. Lapp, *Inorg. Chem.* **1**, 401 (1962).
304. D. A. Buckingham, C. E. Davis, and A. M. Sargeson, *J. Am. Chem. Soc.* **92**, 6159 (1970).
305. M. R. Hoffmann and J. O. Edwards, *Inorg. Nucl. Chem. Lett.* **10**, 837 (1974).
306. J. Scherzer and co-workers, *Inorg. Chem.* **5**, 847 (1966).
307. C. A. Root and J. Williams Allison, *Inorg. Chem.* **9**, 2791 (1970).
308. S. Gabriel and C. Hirsch, *Ber.* **29**, 2747 (1896).
309. J. N. Kukuskin and V. B. Ukrainev, *Z. Neorg. Chim.* **15**, 60 (1970).
310. I. A. G. Roos, A. J. Thomson, and S. Mansy, *J. Am. Chem. Soc.* **96**, 6484 (1974).
311. H. P. Fritz and G. Hierl, *Z. Naturforsch.* **26B**, 476 (1971).
312. N. S. Nametkin and co-workers, *Izv. Akad. Nauk SSSR, Ser. Khim.*, 1874 (1972).
313. V. B. Ukraintsev and Y. N. Kukushkin, *Zh. Obshch. Khim.* **56**, 2292 (1986) and references therein.
314. V. B. Ukraintsev, S. V. Yakovlev, and Y. N. Kukushkin, *Zh. Obshch. Khim.* **59**, 1602 (1989).
315. V. B. Ukraintsev and co-workers, *Zh. Obshch. Khim.* **58**, 389 (1988).
316. V. B. Ukraintsev, S. V. Yakovlev, and Y. N. Kukushkin, *Zh. Obshch. Khim.* **53**, 2066 (1983).
317. W. Beck, W. Danzer, and R. Hüfer, *Angew. Chem.* **85**, 87 (1973).
318. W. Beck and co-workers, *Angew. Chem.* **88**, 511 (1976).
319. W. Danzer and co-workers, *Z. Naturforsch.* **39B**, 167 (1984).
320. P. K. Wong and co-workers, *J. Am. Chem. Soc.* **99**, 2823 (1976).
321. L. S. Hegedus and co-workers, *Inorg. Chem.* **16**, 1887 (1977).
322. H. Alper, F. Urso, and J. H. Smith, *J. Am. Chem. Soc.* **105**, 6737 (1985).
323. H. Alper, *Pure Appl. Chem.* **60**, 35 (1988).
324. M. M. Singh and R. J. Angelici, *Angew. Chem.* **95**, 160 (1983).
325. M. M. Singh and R. J. Angelici, *Angew. Chem. Suppl.*, 184 (1983).
326. M. M. Singh and R. J. Angelici, *Inorg. Chem.* **23**, 2691 (1984).
327. *Ibid.*, p. 2699.
328. R. Bertani and co-workers, *NATO ASI Ser., Ser. C, (Adv. Met. Carbene Chem.)*, **269** 67 (1989).
329. R. A. Michelin and co-workers, *Inorg. Chim. Acta* **174**, 3 (1990).
330. L. Zanotto, R. Bertani, and R. A. Michelin, *Inorg. Chem.* **29**, 3265 (1990).
331. R. Bertani and co-workers, *Inorg. Chim. Acta* **189**, 175 (1991).
332. L. M. Atagi and co-workers, *J. Am. Chem. Soc.* **113**, 870 (1991).
333. **U.S. Pat. 3,242,166** (May 26, 1965), C. G. Krespan and W. J. Middleton (to E. I. du Pont de Nemours & Co., Inc.).
334. C. H. Wang and S. G. Cohen, *J. Org. Chem.* **26**, 3301 (1961).
335. C. L. Stevens and co-workers, *J. Org. Chem.* **29**, 3146 (1964).
336. A. Kubik, K. Piatkowski, and H. Kuczynski, *Roczniki Chemii* **48**, 1225 (1974).
337. J. Werry and co-workers, *Tetrahedron* **45**, 5015 (1989).
338. T. Hiyama, H. Koide, and H. Nozaki, *Bull. Chem. Soc. Jpn.* **48**, 2918 (1975).
339. S. V. Zotova, G. V. Loza, and M. Y. Lukina, *Bull. Acad. Sci. USSR*, 415 (1967).
340. J. Meier, F. Akermann, and H. H. Günthard, *Helv. Chim. Acta* **51**, 1686 (1968).
341. N. R. Bertoniere and G. W. Griffin, in O. L. Chapman, ed., *Organic Photochemistry*, Vol. 3, Marcel Dekker, Inc., New York, 1973, p. 158.
342. S. Braslavsky and J. Hecklen, *Chem. Rev.* **4**, 473 (1977).
343. C. Luner and H. Gesser, *J. Phys. Chem.* **62**, 1148 (1958).
344. R. F. Klemm, *Can. J. Chem.* **43**, 2633 (1965).
345. A. Jones and F. P. Lossing, *Can. J. Chem.* **45**, 1685 (1967).
346. R. F. Klemm, *Can. J. Chem.* **45**, 1693 (1967).
347. M. Kawasaki and co-workers, *J. Chem. Phys.* **59**, 6321 (1973).
348. M. Kawasaki, M. Iwasaki, and I. Tanaka, *J. Chem. Phys.* **59**, 6328 (1973).
349. A. A. Scala and D. Salomon, *J. Chem. Phys.* **65**, 4455 (1976).
350. W. S. Jamieson and C. A. Winkler, *J. Phys. Chem.* **60**, 1542 (1956).
351. B. Bigot, A. Devaquet, and A. Sevin, *J. Org. Chem.* **45**, 97 (1980) and references therein.
352. J. A. Kaye and D. F. Strobel, *Icarus* **54**, 417 (1983).
353. J. P. Ferris and Y. Ishikawa, *J. Am. Chem. Soc.* **110**, 4306 (1988).
354. E. J. Goethals, in K. J. Ivin and T. Saegusa, eds., *Ring-Opening Polymerization*, Vol. 2, Elsevier Applied Science Publishers, New York, 1984, p. 715.
355. D. A. Tomalia and G. R. Killat, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, Vol. 1, John Wiley & Sons, Inc., New York, 1985, p. 680.
356. S. Kobayashi, *Progr. Polym. Sci.* **15**, 751 (1990).
357. E. J. Goethals and R. DeClercq, in W. J. Mijs, ed., *New Methods for Polymer Synthesis*, Plenum Press, New York, 1992, p. 67.
358. T. Saegusa and S. Kobayashi, in Ref. 16, p. 55.
359. K. F. Weyts and E. J. Goethals, *Polymer Bulletin* **19**, 13 (1988).

360. D. A. Tomalia, A. M. Naylor, and W. Goddard III, *Angew. Chem.* **102**, 119 (1990).
361. G. R. Hansen and T. E. Burg, *J. Heterocycl. Chem.* **5**, 305 (1968).
362. S. Tsuboyama and co-workers, *Tetrahedron Lett.*, 1367 (1970).
363. **U.S. Pat. 4,093,615** (June 6, 1978), G. E. Ham and R. L. Krause (to Dow Chemical Co.).
364. H. Wenker, *J. Am. Chem. Soc.* **57**, 2328 (1935).
365. H. Kindler and co-workers, *Chemie-Ing.-Techn.* **37**, 400 (1965).
366. M. Ueshima and Y. Shimasaki, *Kagaku to Kogyo* **43**, 349 (1990).
367. J. Epstein, R. W. Rosenthal, and R. J. Ess, *Anal. Chem.* **27**, 1435 (1955).
368. R. Preussmann, H. Schneider, and F. Epple, *Arzneimittel-Forsch.* **19**, 1059 (1969).
369. D. H. Rosenblatt, P. Hlinka, and J. Epstein, *Anal. Chem.* **27**, 1290 (1955).
370. T. R. Crompton, *Analyst* **90**, 107 (1965).
371. D. J. Evans, R. J. Mayfield, and I. M. Russell, *J. Chromatog.* **115**, 391 (1975).
372. R. Morales, J. F. Stampfer, and R. E. Hermes, *Anal. Chem.* **54**, 1340 (1982).
373. A. M. Thiess, *Arch. Toxikol.* **21**, 67 (1965).
374. *U.S. Code of Federal Regulations: Occupational Safety and Health Standards*, U.S. Government Printing Office, Washington, D.C., 1973, Title 29, part 1910.93, p. 114.
375. Technische Regeln für Gefahrstoffe (TRGS) 102, *Technische Richtkonzentrationen (TRK) für gefährliche Stoffe*, Kuehn, Birett, Bonn, Germany, 1993.
376. *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Man*, Vol. 9, *Some Aziridines, N-, S- and O-Mustards and Selenium*, International Agency for Research on Cancer, Geneva, Switzerland, 1975.
377. L. Verschaeve and M. Kirsch-Volders, *Mutat. Res.* **238**, 39 (1990).
378. S. D. Silver and F. P. McGrath, *J. Ind. Hyg. Toxicol.* **30**, 7 (1948).
379. R. M. V. James, *Biochem. Pharmacol.* **14**, 915 (1965).
380. H. F. Smyth, Jr., J. Seaton, and L. Fischer, *J. Ind. Hyg. Toxicol.* **23**, 259 (1941).
381. C. P. Carpenter, H. F. Smyth, Jr., and C. B. Shaffer, *J. Ind. Hyg. Toxicol.* **30**, 2 (1948).
382. D. Horn and F. Linhart, in J. C. Roberts, ed., *Paper Chemistry*, Blackie, London, 1991, p. 44.
383. D. Horn and J. Melzer, *Das Papier* **29**, 534 (1975).
384. J. Weigl, J. Hlavatsch, and W. Scheid, *Wochenblatt für Papierfabrikation* **13**, 587 (1989).
385. R. Nicke and M. Tappe, *Zellstoff und Papier* **37**, 16 (1988).
386. **U.S. Pat. 3,520,774** (Sept. 28, 1967), H. H. Roth (to Dow Chemical co.).
387. **Eur. Pat. 0 025,515** (Aug. 14, 1980), O. Hertel and co-workers, (to BASF AG).
388. M. Hoffmann, *Zellstoff und Papier* **34**, 134 (1985).
389. F. Linhart and co-workers, *Tappi J.* **70**, 79 (1987).
390. J. G. Penniman, *Paper Trade J.* **162**, 36 (1978).
391. **Rus. Pat. 6 0071-795-A** (Sept. 27, 1983), (to Nippon Shokubai Kagaku Kogyo Co., Ltd.).
392. **Ger. Pat. 3,316,179 A1** (May 4, 1983), F. Reichel and G. Frey (to BASF AG).
393. **Ger. Pat. 3,719,480-A** (Aug. 5, 1986), H.-J. Degen, S. Pfohl, and A. Stange (to BASF AG).
394. L. Westfelt, *Cellul. Chem. Technol.* **13**, 813 (1979).
395. **Eur. Pat. 374,861-A** (Dec. 20, 1988), R. Cohen and G. Luzio (to Hercules Inc.).
396. **Ger. Pat. 2,911,533** (Apr. 10, 1978), E. Litzinger (to British-American Tobacco Co. Ltd.).
397. A. G. Andreopoulos, *Polym. Adv. Technol.* **2**, 87 (1991).
398. **Rus. Pat. 2,225,244-A** (Mar. 27, 1986), S. Yuji and co-workers (to Nippon Shokubai Kagaku Kogyo Co., Ltd.).
399. H. Thiele and K. H. Gronau, *Makromol. Chem.* **59**, 207 (1963).
400. S. Kobayashi and co-workers, *Macromolecules* **20**, 1496 (1987).
401. K. Geckeler, B. Rivas, and R. Zhou, *Angew. Makromol. Ch.* **193** 195 (1991); B. Rivas and K. Geckeler, *Adv. Polym. Sci.* **102**, 171 (1992).
402. Y. Bao and C. Pitt, *J. Polymer Sci., Polymer Chemistry Edition* **28**, 741 (1990).
403. Y. Nakamura and H. Hirai, *Chem. Lett.*, 1997 (1972).
404. J. Anzai and co-workers, *Bl. Chem. Soc. Japan* **56**, 2541 (1983).
405. B. Rivas and co-workers, *Polymer Bull.* **14**, 239 (1985).
406. M. Morita, M. Higuchi, and I. Sakata, *J. Appl. Polymer Sci.* **34**, 1013 (1987).
407. D. Mortimer, *Polymer Int.* **25**, 29 (1991).
408. **Eur. Pat. 61,169** (Mar. 25, 1981), H.-J. Degen and co-workers, (to BASF AG).
409. Y. Kurimura, *Adv. in Polymer Sci.* **90**, 105 (1989).
410. G. Karpuchina and co-workers, *Plasticheskie Massy* **10**, 43 (1988).
411. G. V. Karpuchina and co-workers, *Plasticheskie Massy* **10**, 43 (1988).
412. **Ger. Pat. 3,309,866** (Mar. 19, 1982), N. Kobayashi and A. Funamoto (to Nippon Soda Co. Ltd.).
413. **Ger. Pat. 3,517,107** (May 11, 1985), D. Stoeckigt and co-workers (to BASF AG).
414. **U.S. Pat. 4,415,705** (July 19, 1982), F. Hutter (to Inmont Corp.).
415. **U.S. Pat. 4,165,409** (Feb. 23, 1978), J. Maine.
416. **U.S. Pat. 4,086,385** (Nov. 5, 1975), S. Rowland (U.S., represented by the Secretary of Agriculture).
417. **U.S. Pat. 4,454,254** (Dec. 1, 1982), C. Reichel, J. Patton, Jr., and T. Narayan (to BASF Wyandotte Corp.).
418. Anonymous, *Säddeutsches Kunststoff-Zentrum Würzburg*, Fachtagung, Stuttgart, Germany, 1984, p. 1.
419. W. Marconi, *Reactive Polymers, Ion Exchangers, Sorbents* **11**, 1 (1989).
420. G. Manecke and S. Heydolph, *Makromol. Ch. and Physics, Basel* **182**, 2641 (1981).
421. **Ger. Pat. 2,915,135** (Apr. 19, 1978), M. Gestrelus (to Novo Ind.).
422. **Brit. Pat. 2,082,591** (Aug. 21, 1980), J. Shimizu, K. Suzuki, and Y. Nakajima (to Mitsui Sugar Co. Ltd.).
423. P. Lozano and co-workers, *Process Biochem.* **23**, 75 (1988).
424. **Ger. Pat. 2,828,369** (July 1, 1977), H. Mano (to Sumitomo Electric Ind.).
425. J. Batulin and B. L. Rivas, *J. Makromol. Ch., Rapid Comm.* **2**, 375 (1981).
426. **U.S. Pat. 4,808,215** (Apr. 3, 1987), J. S. Gill and R. G. Varsanik (to Calgon Group).
427. **Ger. Pat. 3,111,713** (Mar. 25, 1981), G. Frey and co-workers, (to BASF AG).
428. **World Pat. 8,201,993** (Dec. 18, 1980), H. Meyer.
429. H. E. Bille and R. R. Fikentscher, *Chem. Aftertreat. Text.*, (1971).
430. N. R. Bertoniere, M. K. Black, and S. P. Rowland, *Textile Res. J.* **48**, 664 (1978).
431. J. R. Cook and D. E. Rivett, *Text. Res. J.* **51**, 596 (1981).
432. C. E. Morris and C. M. Welch, *Text. Res. J.* **53**, 725 (1983).
433. P. E. Ellingham, P. G. Gordon, and J. Rosevaer, *Text. Res. J.* **53**, 577 (1983).
434. **Jpn. Pat. 55 065,234** (Nov. 9, 1978), T. Hiroshi and T. Hidetoshi (to Nippon Shokubai Kagaku Kogyo Co., Ltd.).
435. K. Matsumoto, A. Suganuma, and D. Kunui, *Powder Technol.* **25**, 1 (1980).

436. **Jpn. Pat. 9 133,000** A (Jan. 21, 1983), J. Masa-Mitsu and co-workers (to Nippon Shokubai Kagaku Kogyo Co., Ltd.).
437. V. Valtshev and co-workers, *Zellstoff und Papier* **34**, 88 (1985).
438. A. Naka and co-workers, *Nippon Kagaku Kaishi* **10**, 1342 (1986).
439. G. D. Jones and R. C. Asperger, *J. Macromol. Sci., Chem. J. Macromol. Sci.* **A13**, 835 (1979).
440. L. I. Metelskaja and co-workers, *Ukr. Chim. Z.* **54**, 387 (1988).
441. **Eur. Pat. 561** (July 28, 1977), W. Streit and co-workers, (to BASF AG).
442. **Ger. Pat. 2,934,228** (Aug. 24, 1979), C. C. Heuck, E. Lueddecke, and D. Horn (to BASF AG).
443. A. G. Ludwick and C. G. Overberger, *Rev. Roumaine Chim.* **25**, 1011 (1980).
444. R. E. Wingard, Jr., *Organic Coatings and Appl. Polymer Sci. Proceedings* **44**, 69 (1981).
445. L. Wayne Noell and C. C. Cherry, *J. Med. Chem.* **11**, 63 (1968).
446. C. E. Carraher, Jr. and co-workers, *Organic Coatings and Applied Polymer Sci. Proceedings* **49**, 210 (1983).
447. C. E. Carraher, Jr. and co-workers, *J. Polym. Mat.* **1**, 116 (1984).
448. T. Sato and co-workers, *Polymer Preprints, Japan, Engl. Ed.* **38**, E 357 (1989).
449. R. W. Franck, *Fortschritte. Chem. Org. Naturst.* **38**, 1 (1979).
450. H. Waldmann, *Nachr. Chem. Tech. Lab.* **40**, 235 (1992) and references therein.
451. **U.S. Pat. 5,032,278** (Aug. 6, 1990), J. Reale, Jr. (to Texaco Inc.).
452. T. Sata, *Pure Appl. Chem.* **58**, 1613 (1986).
453. J. E. Cadotte and co-workers, *J. Makromol. Sci., Part A: Chemistry* **A15**, 727 (1981).
454. Y. Maeda and co-workers, *Polymer J.* **23**, 501 (1991).
455. M. F. Tsai and M. Levy, *J. Polymer Sci., Polym. Chem. Ed.* **22**, 2523 (1984).
456. K. Uehara and K. Kawai, *Polymer Preprints, Japan, Engl. Ed.* **38**, E 551 (1989).
457. T. Sata and R. Iuzo, *Angew. Macrom. Chem.* **171**, 101 (1989).
458. L. E. Bromberg, *J. Member. Sci.* **62**, 131 (1991).
459. **Eur. Pat. 147,743** (Dec. 14, 1984), R. Baur, K. Oppenlander, and K. Barthold (to BASF AG).
460. **Eur. Pat. 444 542** (Mar. 2, 1990), K. Oppenlander and co-workers (BASF Corp.).
461. **Ger. Pat. 3,712,317** (Apr. 11, 1987), H.-P. Seelmann-Eggebert and co-workers (to BASF AG).
462. K. Takemoto and Y. Inaki, *Macromol. Sci., Part A: Chem* **A25**, 5–7, 757 (1988).
463. **Eur. Pat. 174,644 A** (Sept. 11, 1985), T. Teraji and co-workers (to Fujisawa Pharmaceutical Co., Ltd.).
464. L. W. Woods, A. B. Borkevec, and F. M. Harz, *J. Med. Chem.* **7**, 371 (1964).

G. Scherr
U. Steuerle
R. Fikentscher
BASF AG

IMMUNOASSAY

Immunoassay is a method that identifies and quantifies unknown analytes using antibody–antigen reactions. Techniques are based in immunochemistry, analytical chemistry, and biochemistry, with a history of development paralleling advances in microbiology and immunology (see also IMMUNOTHERAPEUTIC AGENTS).

Immunoassay has developed primarily as a subdiscipline of clinical diagnostics. From the first agglutination assays through radioimmunoassay (RIA) and other immunoassay formats, the focus of each advancement was a specific clinical need, such as the identification of infectious agents in biological tissues and fluids. The development and commercialization of RIAs in the 1960s and 1970s placed immunoassay as a leading technology in clinical diagnostics, eventually evolving into a variety of immunoassay formats which, in 1992, captured over 25% of the *in vitro* clinical diagnostic market (1).

The success of immunoassay in clinical diagnostics, and the generic nature of immunoassay technology has resulted in the application of the method in other areas. By the late 1980s, commercial immunoassay products and systems were available for detection and diagnosis in environmental, food, and chemical processing applications. Whereas the application of immunoassays in these areas is small in relation to clinical diagnostic immunoassays, nonclinical applications of immunoassays hold the potential for significant growth by the year 2000 (see AUTOMATED INSTRUMENTATION, CLINICAL CHEMISTRY).

Historical Perspective

The modern discipline of immunochemistry began in the late 1800s when Ehrlich and others discovered specific serum agents, now known as antibodies, that neutralized infectious agents (2). Based on the discoveries of specific precipitation of plague bacilli with serum from animals previously exposed to the bacillus and specific agglutination clumping of bacterial cells (3), one of the first diagnostic immunoassays, an agglutination test for typhoid, was developed. By the 1930s, precipitating antibodies were being used routinely to quantify bacteria. For example, Heidelberger and Kendall developed the typhoid precipitation test into a quantitative diagnostic method for typhoid, able to detect <0.1 μg of bacterial antigen (4).

Immunodiffusion and immunoprecipitation, developed in the 1940s as a means to identify and semiquantitate specific proteins, were the direct precursors to the development in 1953 of immunoelectrophoresis, a method used in many clinical laboratories (5). Single- and double-gel immunodiffusion and immunoelectrophoresis were, in effect, the first standardized and routinely used immunoassay methods (see ELECTROSEPARATIONS, ELECTROPHORESIS).

Although useful, immunodiffusion and immunoelectrophoresis are tedious and time consuming, difficult to automate, and not easily applied to mass sample analysis. In the early 1960s, a method was introduced that revolutionized immunoassay and clinical diagnostics. Developed from the pioneering work of Berson and Yalow (6), radioimmunoassay became a routine, automated, highly adaptable and cost-effective method of analyzing large numbers of clinical samples. Quickly commercialized, by the early 1970s hundreds of RIAs were available for not only a wide spectrum of clinical analytes, but also environmental, food, and chemical analytes.

The success of RIA led to the development of other immunoassay methods in the late 1960s and early 1970s. Whereas these methods used the basic format of RIAs, the quantifiable indicators were changed from radioactive isotopes (qv) to ones based on the production of color, fluorescence, or light. The first of the nonisotopic immunoassays to be developed and commercialized utilized enzymes as indicator agents and utilized technology based on the work of Avrameas for enzymatic labeling of antibodies and antigens (7). These assays became known as enzyme-linked immunoadsorbant (or immunosorbant) assays (ELISAs) or simply enzyme immunoassay (EIA), and have become the primary laboratory and commercially used immunoassay method, applicable to hundreds of analytes and available in kit or automated analyzer format. EIAs and other nonradioactive immunoassays have largely replaced RIAs. Further evolution of immunoassays includes the development of near real-time, homogenous immunoassays, and the adaptation of immunoassay formats to immunosensors, biosensors (qv) based on antigen–antibody reactions.

The Antibody–Antigen Reaction

Immunoassays are based on the binding and complexing of an antigen to an antibody, and the use of some physical or chemical means to measure and quantify the antigen–antibody complex. The antibody–antigen reaction is a typical reversible bimolecular reaction having rate constants for the forward and backward reactions which are dependent on concentration of the antigen (Ag) and antibody (Ab), affinity for the antigen as defined by the association constant of the antibody for its antigen, temperature, pH, and other environmental conditions. This reaction is represented by an equation common to reversible receptor-ligand assays:



and the equilibrium constant for the reaction is determined by the mass action equation:

$$K = \frac{[AbAg]}{[Ab][Ag]}$$

(2)

where [Ab] is the concentration of antibody sites for antigen, and [Ag] is the concentration of free antigen. The association, or affinity, constant for the antibody–antigen reaction is then further defined as the equilibrium constant at half-saturation of the antibody with antigen. Because at half-saturation AbAg and Ab are equal, these cancel in the above equation and the association constant, *K_a*, is equal to the reciprocal of the free antigen concentration:

$$K_a = \frac{1}{[Ag]}$$

(3)

Thus, if the antibody has a high affinity for the antigen, it has a high association constant. Typical association constants range from 10⁸ to 10¹⁰ L mol⁻¹, and as high as 10¹³ L mol⁻¹ for some monoclonal antibodies.

The definition of an association constant for an antibody–antigen reaction can become more complex if the antibody–antigen reaction involves a multivalent antigen, as is the case when a polyclonal antiserum is used for detection of an antigen. This type of multivalent binding is termed avidity and is defined by the equation:



Definition of the association (or avidity) constant for such multivalent antibody–antigen reactions must consider not only the heterogeneity of the antibodies and the antigen determinant site(s), but also an apparent additive effect of binding two antigen molecules to a single antibody. Such effects lead

to a multiplying of the individual association constants and an apparent large increase in the total association–avidity constant. This multiplication of avidity through multivalent binding has been exploited to increase the sensitivity of many immunoassays. A more detailed discussion of antigen–antibody reaction kinetics may be found in the literature (8,9).

As the result of many studies on the antigen–antibody reaction, it is known that the primary antibody–antigen binding event (in solution) occurs in about 10 seconds whereas formation of the final complex (final binding equilibrium) takes about 30 minutes. Antigen–antibody reactions on solid surfaces, such as in the case where the antigen or antibody is chemically immobilized onto a solid support, can take significantly longer, up to a few hours, to reach final binding equilibrium. This time dependence of antigen–antibody binding plays a central role in immunoassay development and the need to balance maximal antigen–antibody binding within a minimal amount of time.

Most antibody–antigen reactions are optimal at around pH 7 in a low salt medium between 0 and 40°C. The typical free energy changes associated with the antibody–antigen reaction are around 33 to 42 kJ mol⁻¹ (8 to 10 kcal mol⁻¹). This is about the same amount of energy change associated with the nonspecific binding of drugs to a plasma protein such as serum albumin (see BLOOD FRACTIONATION). The free energy varies depending on the nature of the antigen, being the lowest for large antigens, such as proteins (qv), and highest for antigens such as small peptides and low molecular weight drugs. Free energy changes are directly related to antibody affinity for the antigen. The free energy of the antigen–antibody binding event and antibody affinity for its antigen are thus also important factors to be considered during the development of an immunoassay.

Basic Technology

The principle approach to immunoassay is illustrated in Figure 1, which shows a basic sandwich immunoassay. In this type of assay, an antibody to the analyte to be measured is immobilized onto a solid surface, such as a bead or a plastic (microtiter) plate. The test sample suspected of containing the analyte is mixed with the antibody beads or placed in the plastic plate, resulting in the formation of the antibody–analyte complex. A second antibody which carries an indicator reagent is then added to the mixture. This indicator may be a radioisotope, for RIA; an enzyme, for EIA; or a fluorophore, for fluorescence immunoassay (FIA). The antibody–indicator binds to the first antibody–analyte complex, free second antibody–indicator is washed away, and the two–antibody–analyte complex is quantified using a method compatible with the indicator reagent, such as quantifying radioactivity or enzyme–mediated color formation (see AUTOMATED INSTRUMENTATION, CLINICAL CHEMISTRY).

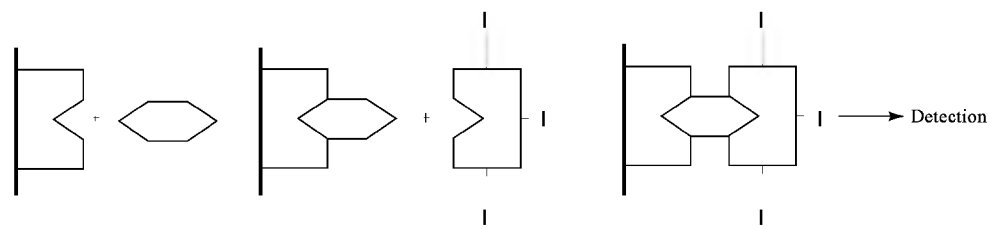


Fig. 1. A principal approach to immunoassay, the sandwich immunoassay, where the thick line represents the solid matrix,  the antibody,  the antigen, and I an indicator molecule such as an enzyme, fluorophore, or radioisotope.

In fact, most RIAs and many nonisotopic immunoassays use a competitive binding format (see Fig. 2). In this approach, the analyte in the sample to be measured competes with a known amount of added analyte that has been labeled with an indicator that binds to the immobilized antibody. After reaction, the free analyte–analyte–indicator solution is washed away from the solid phase. The analyte–indicator on the solid phase or remaining in the wash solution is then used to quantify the amount of analyte present in the sample as measured against a control assay using only an analyte–indicator. This is done by quantifying the analyte–indicator using the method appropriate for the assay, for example, enzyme activity, fluorescence, radioactivity, etc.

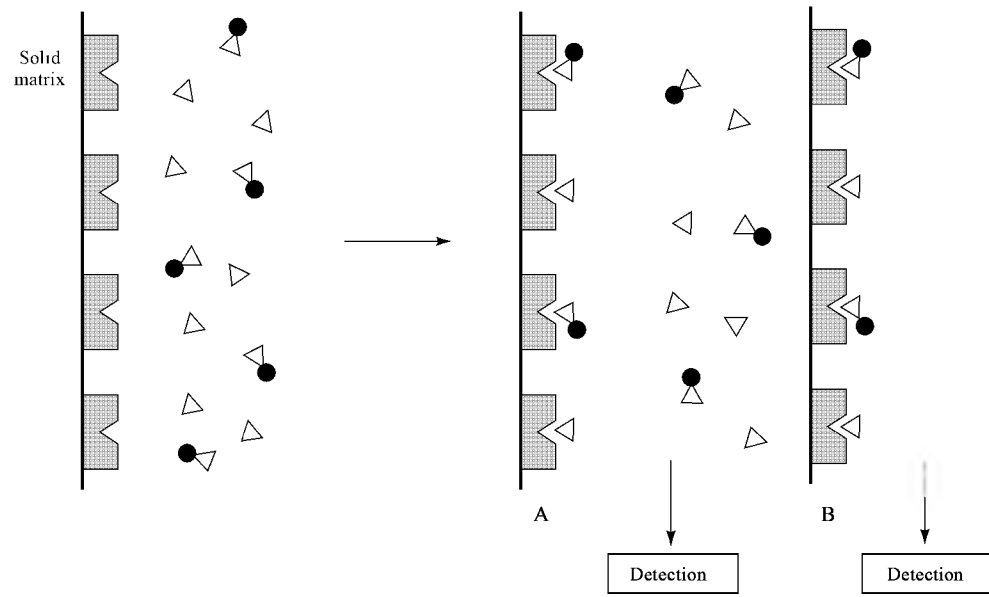


Fig. 2. The basic approach for a competitive immunoassay. The analyte () and analyte–indicator () compete for sites on the antibody () which may be immobilized or in solution. Quantitation may utilize the remaining (free) indicator in solution A, or the indicator bound to the antibody, B.

There are many variations on these two basic approaches for immunoassays (Figs. 1 and 2). For example, antigen instead of antibody may be immobilized onto a solid surface to allow antibody detection in samples. A displacement rather than a competitive reaction may be used, ie, the analyte displacing analyte–indicator bound to the antibody. Additionally, supports can be designed to allow better separation and quantification. For example, the capture antibody (or antigen) may be immobilized onto magnetic beads. After the immunoreaction, the antibody–analyte complex on the beads can be separated rapidly using a magnetic field.

Immunoassay Design. The basic reagent and design requirements of an immunoassay are antibody, antigen, conjugates of either or both, and a means for separating bound and unbound reagents. Both antibodies and antigens may be used per se in an assay or as indicator conjugates. Conjugating involves the chemical linkage of the antibody or antigen to another molecule such as a radioactive isotope; an enzyme, usually peroxidase,

alkaline phosphatase, or glucose oxidase [9001-37-0]; a fluorophore, usually fluorescein [2321-07-5] or rhodamine B [81-88-9]; or a chemiluminescent molecule, such as luminol [521-31-3]. Conjugation reactions are discussed in detail elsewhere (7,10–12). In summary, the antibody or antigen is linked to the labeling molecule using a variety of chemical methods including glutaraldehyde cross-linking; carbodiimide, epichlorohydrin, and N-hydroxysuccinimide coupling; mixed anhydride formation and coupling; and periodate cleavage followed by reductive alkylation. In most cases, the conjugation reactions occur spontaneously in a buffered solution and the products–reactants can be easily separated by chromatography (qv) or ultrafiltration (qv).

Noncovalent methods for conjugating antibodies and antigens are also used in many immunoassays. For example, one well-documented approach labels the antigen and or antibody using either biotin [58-85-5] or streptavidin, and then uses the conjugates in a variety of competitive and displacement type immunoassays. The advantage in the use of such conjugates is the extremely high affinity constant of the avidin–biotin complex, estimated at approximately $\frac{10^{14}}{\text{L mol}}$ (13).

Once the appropriate reagents are prepared for an immunoassay, a laboratory assay is developed to serve both as a means of assessing the performance of the individual assay components and as a prototype for the final assay. The prototype is then used for reagent optimization and development of the final assay methodology. In the case of commercial immunoassays for clinical applications, development from an initial laboratory prototype of the assay, which includes development of antibodies and reagents, through final market release may take from 6 to 24 months depending on whether the new assay is an extension of existing, in-place technology, or represents an entirely new assay format and or application (1).

There are many possible means for quantification of the antigen–antibody reaction. Immunoassays may be classified according to the technology used for detection and quantification of the analyte being detected.

Turbidimetric Agglutination Immunoassays. Agglutination–precipitation immunoassays were among the first practical applications of the antigen–antibody reaction in diagnostic tests. These assays are not as widely used in the 1990s as EIA and FIA because they are either not quantitative enough or lack the sensitivity limits of RIA, EIA, and FIA.

Classic agglutination assays utilized the formation and visual determination of antibody–antigen precipitates resulting from the formation of insoluble complexes between antibodies and high molecular weight antigens, or between antibodies and cell wall or surface antigens of microbes or animal cells. These simple precipitation reactions also form the basis for immunoelectrophoresis which is the quantification of antigenic determinants of a cell after electrophoretic separation in a gel by the addition of specific antibody and precipitation in the gel.

Immunoprecipitation is also the basis for immunodiffusion methods such as Ouchterlony double immunodiffusion (14). The most frequently used method for this procedure is to apply various antigen (or antibody) solutions to wells punched out of an agar base which surrounds a central well containing antibody (or antigen). The antigen–antibodies diffuse through the gel and when an antibody–antigen reaction occurs, a distinct precipitin line results. Variations of well locations and contents can result in the identification and semiquantification of an antigen (or antibody).

Latex particles, in most cases actually made of monomeric polystyrene, coated with a specific antibody or antigen are the basis for a number of commercial agglutination assays (see LATEX TECHNOLOGY). Particle reaction with the target antigen or antibody in a sample results in antigen–antibody binding and, if enough target analyte is present, the clumping or agglutination of the particles as the antigens or antibodies bridge two or more particles to form a visible precipitate.

Latex agglutination immunoassays are easily formatted into simple kits which can provide yes–no and semiquantitative estimates of antigen (or antibody) in a sample. The first such assay was developed in 1957 for rheumatoid factor (15) and assays are on the market for the determination of many species of bacteria, fungi, Mycoplasma, parasites, *Rickettsia*, and viruses, as well as for the determination of autoimmune disease, hormones (qv), drugs (see PHARMACEUTICALS), and blood proteins (16). Latex agglutination is also the basis of many home pregnancy tests.

An example of a latex bead-based diagnostic for Group A Streptococcal antigen (17) manufactured and sold for *in vitro* use in physician's offices and clinical laboratories is shown in Figure 3. In fact, the assay is based on immunochromatography, a combination of immunoassay and chromatography. The chromatography is used for the separation of bound and free antibody and or antigen. Initially, a specific carbohydrate containing antigen is extracted from bacteria collected on a throat swab sample from the patient. The extract is then added to the beginning of a paper strip in a plastic housing about the size of a tape cassette. There the antigen (if present) reacts with antibody to the Streptococcal antigen which has been immobilized onto a blue-colored latex bead. The solution then migrates on the strip, carrying the latex bead–antibody–antigen complex with it until it reaches a location where a second antibody to the Streptococcal antigen has been chemically immobilized. If the antigen is present, a sandwich of immobilized antibody–antigen–antibody-bead forms and a blue line, indicative of a positive result, forms. The assay takes approximately five minutes and has a reported sensitivity of >96%, ie, <4% false positives or negatives, and a specificity of >97% tested against 40 specific bacteria including 16 Streptococcal species (18). The sensitivity of the test appears to be approximately 10^8 to 10^9 organisms.

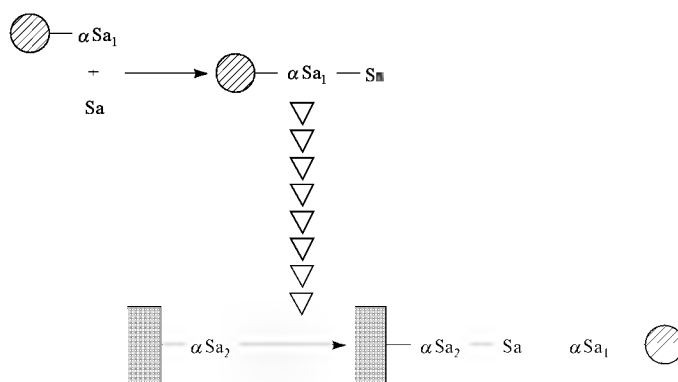


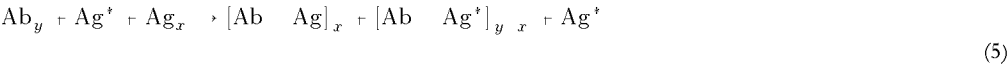
Fig. 3. Example of an immunochromatographic assay for a Streptococcal antigen (Sa) using antibody to the antigen (αSa_1) linked to a blue-colored latex bead (⊙). The downward pointing triangles represent chromatographic migration. Formation of a sandwich linking the antigen between the latex bead and a second, immobilized antibody (αSa_2) results in an immobilized colored complex giving a positive visual test (18).

Latex agglutination tests can also be quantitated using a particle counter, spectrophotometer, or nephelometer. These detectors can overcome the sensitivity problems associated with visual determination of the agglutination reaction, increasing the sensitivity for a protein, for example, from the ppm or $\mu\text{g mL}$ to the ppb or ng mL range. Such assays and support equipment are offered by a number of commercial companies for blood factors and proteins, eg, rheumatoid factor, α -fetoprotein, plasminogen [9001-91-6], prothrombin [9001-26-7], and fibrinogen [9001-32-5]; and drugs, eg, gentamicin [1403-66-3], phenobarbital [50-06-6], and theophylline [58-55-9].

Radioimmunoassay. RIA caused a revolution in clinical diagnostics, providing a rapid, ultrasensitive method for the detection of nearly any agent to which an antibody could be developed. Sensitivity limits in the parts-per-trillion (ppt) or pg mL are commonplace and RIAs are available for nearly every class of clinically relevant analyte including drugs, hormones, immunoglobulins, immune complexes, blood factors, microbes, viruses, and tumor antigens.

Most modern RIAs utilize a competitive assay format (Fig. 2) in which radiolabeled antigen, Ag^* , competes with unlabeled antigen, Ag, in a sample for binding to the antibody, Ab. The free antigens are then separated from the antigen–antibody complexes, and the amount of radioactivity in the

complex, or that remaining free in the wash, is quantified by radioactivity counting and related to antigen concentration. This reaction may be represented by



and assumes that $[\text{Ag}^*] > [\text{Ab}] > [\text{Ag}]$. In this case, the antibody–antigen complexes could be captured for counting after removal of the free Ag^* from the reaction by ion exchange or gel permeation chromatography; adsorption of the complex onto charcoal, plastic, or a filter; or by addition of another antibody to result in immunoprecipitation.

The primary limitation to RIA is its need for development, purification, and standardization of radioactive reagent(s). Moreover methods must be in place for the safe use and disposal of the assay reagents. One of the most commonly used radioisotopes for RIAs is iodine-125 $[14158-31-7]^{125}\text{I}$. This isotope has a half-life of 60 days, requiring frequent resynthesis and purification of assay reagent(s). Many proteins labeled with I^{125} also undergo significant autodegradation in as little as two weeks owing to the incorporated radioactive iodine.

In spite of these drawbacks, RIA remains a principal immunoassay method and it is expected to continue to be used extensively in many clinical and research laboratories for the foreseeable future.

Enzyme Immunoassay. In EIA, antibody (or antigen) is labeled with (or conjugated to) an enzyme, and this reagent is used to complex and quantify the target antigen (or antibody) in a sample. Conjugation may utilize a variety of chemical methods.

The primary use of EIA when it was first developed was for histological labeling and localization of specific cell macromolecules. For example, enzymes labeled with peroxidase were used to locate specific cellular compartments and structures for microscopic examination. The flexibility of EIA was recognized quickly and it was adapted for use as a laboratory assay.

The specific enzyme to be used in an EIA is determined according to a number of parameters including enzyme activity and stability (before, during, and after conjugation), cost and availability of the enzyme substrate, and the desired end point of the EIA, such as color. Most EIAs utilize a colored end point which can be readily determined both visually and spectrophotometrically. Table 1 lists a number of enzymes which have been used in immunoassays and their substrates.

Table 1. Examples of Enzymes Used in EIA

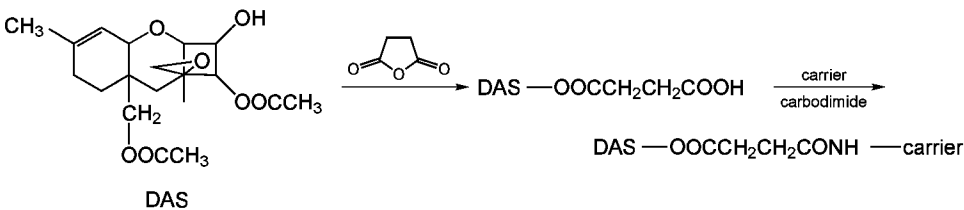
Enzyme	Enzyme classification number	Substrate	Wavelength of detection, nm
aldolase	EC 4.1.2.13	fructosediphosphate	600
alkaline phosphatase	EC 3.1.3.1	p-nitrophenyl-phosphate	405
acetylcholinesterase	EC 3.1.1.7	5,5'-dithiobis(2-nitrobenzoic acid)	412
amylase	EC 3.2.1.1	starch–iodine	440–500
cholesterol oxidase	EC 1.1.3.6	cholesterol–amino-antipyrene	500
creatine phosphokinase	EC 2.7.3.2	α-naphthol–creatine phosphate–diacetyl	520
elastase	EC 3.4.21.36	elastin–congo red	490
glucose oxidase	EC 1.1.3.4	thiozoly blue–phenazine methosulfate	540
β-glucuronidase	EC 3.2.1.31	phenophthalein glucuronide	550
peroxidase	EC 1.11.1.7	o-phenylenediamine, o-dianisidine, 5-aminosalicycliate	492 460 474
urease	EC 3.5.1.5	bromcresol purple	588

EIAs can be used per se or with a spectrophotometer. Traditionally, EIAs have been developed in 96-well microtiter plates which provide the immobilization support for the assay, the reaction vessel, and, when linked to a spectrophotometer-based reader, a rapid means to detect and quantify the color resulting from interaction of a substrate with the antibody–antigen–enzyme complex. Automated immunoassay analyzers targeted primarily for use in the clinical laboratory have taken automation one step further, utilizing robotics to carry out all reagent additions, washings, and final quantification including report preparation.

In addition to instrument-dependent EIAs, a number of EIA kits have been developed for remote lab or field use which depend either on visual determination of color or portable spectrophotometers readers for quantifying the color response. Whereas such field assays are moderately complex (multistep) and require technical training, they are providing rapid means for determination of a variety of analytes, from microbes in food processing (qv) to environmental pollutants. These include commercial assays for herbicides (qv), pesticides, aflatoxins, polychlorinated biphenyls (PCBs), dioxin, and various volatile organic compounds (solvents and chlorinated hydrocarbons), and microorganisms including *Salmonella*, *Listeria*, *Clostridium*, and *Staphylococcus* species. These kits are sensitive for their target analytes in the ppm to ppb range, produce results in 15 to 30 min, and cost from \$1 to \$10 per sample (1). The same technology is being applied to plant and animal diagnostics, including EIA tests for plant viruses, animal diseases, antibiotics in meat and milk, and animal hormones.

There are several basic steps in developing an EIA. One example is the immunoassay developed for the determination of a trichothecene mycotoxin, diacetoxyscirpenol (DAS), also known as anguidine, which is a scirpen-type sesquiterpenoid secondary metabolite of the parasitic plant fungus *Fusarium sambucinum*. DAS appears to cause vomiting when fungus-infected wheat and corn is ingested. DAS has also been shown to have anticancer activity. DAS is an example of a low molecular weight, environmental analyte, and as such, presents different challenges to EIA development as compared to EIAs for high molecular weight analytes.

As a low molecular weight compound DAS is not significantly immunogenic, ie, it is a hapten and thus requires conjugation to a suitable antigenic carrier in order to elicit a successful antibody response in animals. DAS treated with succinic anhydride results in the DAS–hemisuccinate (DAS–HMS) shown.



The available free carboxyl groups of the DAS–HMS can be linked via a peptide bond to available primary amine groups onto highly antigenic carriers using a carbodiimide (19). The carriers used in this case were bovine serum albumin (BSA) and poly-L-lysine (molecular weight 150,000 to 300,000). The

resulting conjugates may be purified using Sephadex G-25 gel permeation chromatography (gpc). Typical DAS:carrier ratios were found to be approximately 25:1 for BSA, and 100 to 350:1 for poly-L-lysine as determined by high pressure-gpc analysis.

Polyclonal antibodies to the DAS conjugates were produced by inoculating mice with conjugate equivalent to 10 to 30 µg of DAS in Freund's complete adjuvant, ip, five times over a four-week period. Ascites fluid was collected at 6 and 10 weeks and processed for antibody. DAS antibody (αDAS) was purified from the ascites fluid, ie, cellular fluid containing antibody which accumulates in the peritoneal cavity of animals inoculated with an immunogenic compound, using the standard purification scheme shown in Figure 4. As is common for polyclonal antibodies raised to a conjugated antigen, further purification was necessary to separate αDAS antibodies from antibodies to the carrier BSA or poly-L-lysine. Using affinity chromatography, in which the original carrier, BSA or poly-L-lysine, is immobilized on Sepharose 4B support material, the antibodies for carrier were adsorbed by the column; antibodies having primarily DAS activity rapidly moved through the column, resulting in αDAS preparations which had a specificity ratio of DAS to carrier of >100 : 1.

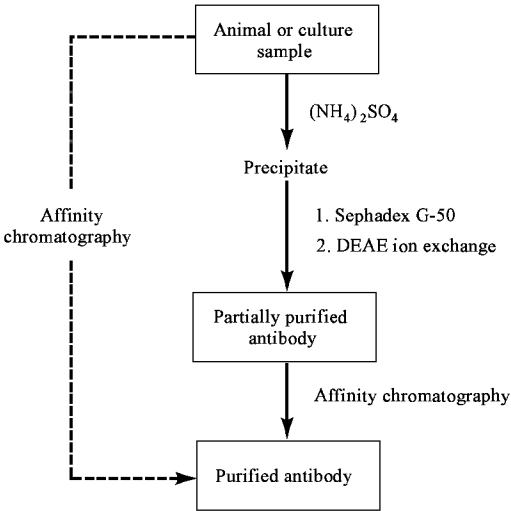


Fig. 4. Example of antibody purification from animal or culture sources. In some cases, affinity chromatography may be used directly with the source material, bypassing the precipitation and other column chromatography steps.

The αDAS antibodies were used as the basis to develop a number of sandwich and competitive EIAs for DAS. In the simplest application, αDAS was conjugated to alkaline phosphatase (AP) using glutaraldehyde as the conjugating agent. After purification by gpc, the αDAS–AP was characterized for DAS specificity using a standard microtiter format, ie, varying amounts of DAS were bound to the microtiter wells and then a constant amount of αDAS–AP was added to all wells. After reacting for 30 to 90 min, free αDAS–AP was washed out of the wells and *p*-nitrophenyl phosphate, an indicator reagent, was added. The production of color in the wells after 15 min was then quantified using a standard microtiter plate reader. Using this approach, one preparation of αDAS–AP was calculated as having a $K_a = 1.9 \times 10^6 \text{ L mol}$ for DAS, as shown in Figure 5.

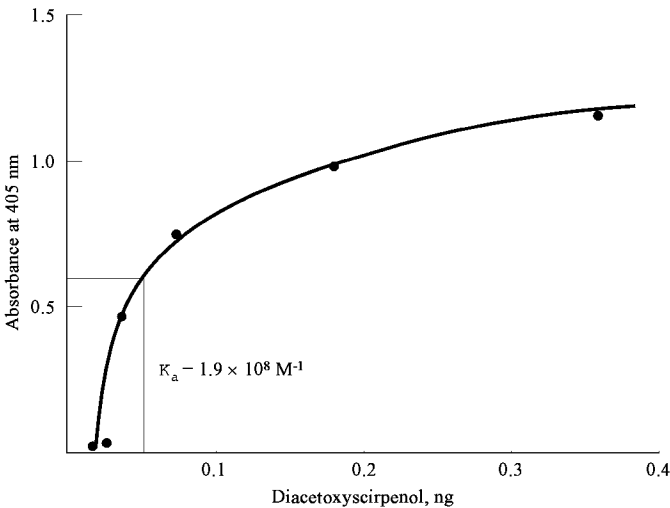


Fig. 5. Determination of the K_a for the binding of DAS to its polyclonal antibody raised in mice. A fixed amount of immobilized antibody in a microtiter plate is reacted with increasing amounts of DAS and the amount of DAS bound at each concentration is determined using an EIA based on alkaline phosphatase (AP) (19).

This same experimental approach can be used to determine the applicability of the αDAS–AP to a competitive assay for DAS. As shown in Figure 6, increasing amounts of free DAS were used to define the 50% inhibition level (ID_{50}) of DAS for binding of two αDAS–AP conjugates to immobilized DAS. This approach was also used to determine the sensitivity of an EIA, as well as the specificity of the assay, as shown in Table 2. Increasing amounts of trichothecene mycotoxins closely related to DAS were added to microtiter plate wells containing a constant amount of prereacted DAS–αDAS–AP. After 30 min, excess toxin and any free toxin–αDAS–AP were washed out, and substrate was added. Quantification of the color produced was directly related to the ability of the added toxin to displace αDAS–AP from the immobilized DAS, which is an indication that the αDAS also has an avidity for that toxin. As shown in Table 2, free DAS, as expected, is its own best displacing agent, whereas only DAS–HMS showed any appreciable displacing capability. This can be expected because the hemisuccinate linker is also immunogenic and leads to the production of antibodies specific for the linker in the polyclonal antibody population. All the other toxins had at least 100× less the avidity for the antibody, illustrating the specificity of the αDAS for DAS.

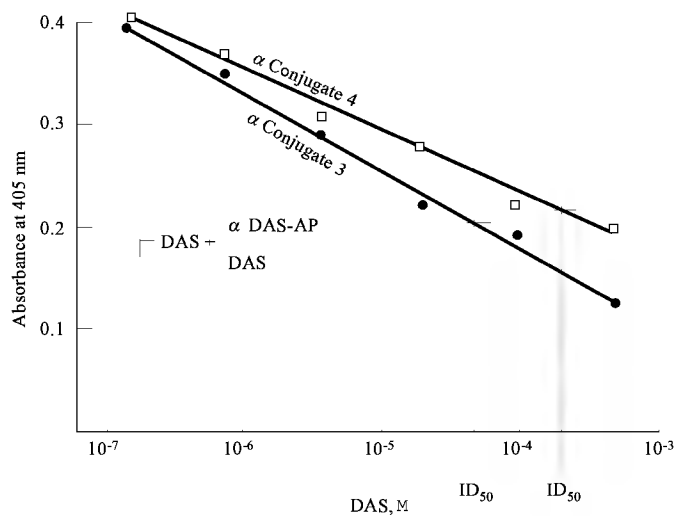


Fig. 6. Determination of the amount of free DAS required to cause 50% binding inhibition, ID₅₀, of αDAS–AP to immobilized DAS, as a means to determine the specificity of an antibody for its target analyte (19).

Table 2. Inhibition of DAS–αDAS Binding by Various Mycotoxins

Mycotoxin	ID ₅₀ ^a	
	μg	10 ^{−6} M
DAS diacetoxyscirpenol	0.21	5.7
DAS–HMS hemisuccinate	16.3	350
T2 toxin	>23	>500
HT2 toxin	>21	>500
verrucarol	>13	>500
verrucarin A	>25	>500
zearaleone	>16	>500

^a The concentration required to displace 50% of αDAS–AP bound to immobilized DAS.

Based on such characterization studies, both sandwich (Fig. 1) and competitive (Fig. 2) EIAs were developed for DAS having a lower detection limit for DAS of 10 to 20 ppb, ie, 10 to 20 ng mL (19).

There are many variations to this basic type of EIA. These include variations aimed at enhancing or increasing the sensitivity and specificity of an assay. For example, an EIA may utilize enzyme amplification to increase the speed and sensitivity of an immunoassay. In this approach, the enzyme label in the EIA produces a substance which triggers a second enzyme-based system which can generate large quantities of color in a very short time. Thus the product of the enzymatic activity of the antigen–antibody–enzyme complex does not need to be detected; rather it can serve as a catalyst to begin the second reaction. The second enzyme system can be present in relatively large quantities, facilitating rapid color formation, because the second enzyme is silent and noninteractive with the assay until the first reaction product turns it on. For example, a standard microtiter plate-based EIA for thyroid stimulating hormone (TSH), using alkaline phosphatase as the enzyme, can be amplified by adding nicotine adenine dinucleotide phosphate (NADP⁺) and a second enzyme system such as alcohol dehydrogenase and lipoamide dehydrogenase, which requires NAD⁺, produced by alkaline phosphatase-dephosphorylation of NADP⁺, for the production of the colored formazan dye (20).

The enzyme mercuripapain, HgP, is being utilized as an enzyme amplification system for immunoassay. HgP is papain (EC 3.4.22.2) which is allosterically inhibited by mercury. Upon removal of the mercury by, for example, a mercaptan, the protease is activated. Displacement EIAs are being developed for the detection of viruses where the virus or a viral antigen is labeled with a mercaptan such as 3-mercaptopropionic acid and bound to immobilized antibody to the virus. Upon addition of a sample containing the virus, the mercaptan-labeled antigen is displaced and is able to sequester mercury away from the HgP present in the reaction mixture. This then activates the enzyme which acts on a suitable substrate such as *N*-α-benzoyl-L-arginine ethyl ester (BAEE), rapidly producing hydrogen ions. The hydrogen ions can be detected directly by use of an electrode or can be linked to an indicator. As of this writing, studies on this system are incomplete but indicate that the system can act as an efficient means to amplify EIAs.

Another approach to enzyme activation allows the EIA to be carried out in a homogenous format in which no washing steps are required. Termed prosthetic group label immunoassay (PGLIA), the method involves the activation of apo-glucose oxidase (apo-GOD) by its cofactor, flavin adenine dinucleotide (FAD). Essentially, FAD is covalently linked to antigen and the labeled antigen competes with sample antigen for antibody. FAD-antigen not bound by antibody can combine with apo-GOD, activate the enzyme, and a color reaction results. The advantage to this approach is that no separation steps are required for the assay. This approach has been used for the determination of drugs such as theophylline (21). A similar system for the detection of gentamicin utilized NAD⁺-labeled drug linked to a NAD⁺ peroxidase glucose-6-phosphate dehydrogenase enzyme system (22).

Fluorescence Immunoassay. Basic FIA follows the same formats and approaches as EIA. The difference lies in the indicator: a fluorophore is used instead of an enzyme. This allows direct quantification of the indicator–antibody–antigen complex, or free indicator–reagent, without the need for a substrate.

FIA was originally developed as a histological technique to localize specific cellular sites using the specificity of an immunological reaction (23). The resulting fluorescent antibodies can be detected in animal tissues at levels as low as 1 μg mL of body fluid. Fluorophore-labeled antibodies have also been used widely for flow cytometry applications using fluorescein antibodies to cell surface markers to detect and quantify specific cells (24).

The most used FIA reagents conjugate a fluorophore such as fluorescein–isothiocyanate (FITC) or rhodamine–isothiocyanate to antibody (or antigen) free amino groups. Examples of other commonly used fluorophores for FIA and their spectral characteristics are presented in Table 3. FIA assays are available in sandwich and competitive formats similar to EIAs. Unlike EIA kits which can be used directly with visual color determination, FIAs require a fluorometer, and thus are primarily laboratory-based.

Table 3. Commonly Used Fluorophores in FIA

Fluorophore	CAS Registry Number	Excitation wavelength ^a	Emission wavelength
-------------	---------------------	------------------------------------	---------------------

Allophycocyanine		620	660–665
7-amino-4-methylcoumarin-3-acetic acid		347	456
bis-benzimide (Hoechst 33258)	[2349-45-4]	340–360	470–480
fluorescein-isothiocuprate FITC	[3326-32-7]	495	525
indocarbocyanine (Cy3)		550	570
indodicarbocyanine (Cy5)		650	680
lissamine rhodamine B-sulfonyl hydrazine		556	580
lucifer yellow CH	[67769-47-5]	428	540
phycocyanine	[11010-15-2]	620	655
phycoerythrin		495	578
rhodamine isothiocyanate	[36877-69-7]	540–550	570
Texas red	[82354-19-6]	596	620

^a Excitation and emission wavelengths are for the unconjugated fluorophore. Wavelengths for conjugates may vary according to the composition of the conjugate.

The high sensitivity achievable with FIA has led to the development of a number of FIA assay formats which attempt to both increase assay sensitivity while decreasing background interference, a primary drawback of FIA. For example, time-resolved fluorimetry (25) is used in one commercial automated FIA system to decrease background fluorescence interference and, as a result, increase sensitivity. The antibody–antigen complex formed in this system includes a second antibody conjugated to europium. When an organic enhancer such as naphthoyltrifluoroacetone is added to the antibody–antigen–antibody–europium complex, a europium-enhancer chelation complex results. Lanthanide chelates have fluorescent lifetimes up to 10 times longer than other fluorophores, such as those occurring naturally and representing fluorescence background (see LANTHANIDES). Therefore measurements of fluorescent signals from the complex can be delayed until most or all of the background fluorescence has decayed, resulting in time-resolved fluorimetry.

Another variation of FIA is based on fluorescence polarization (FP). Based on the random rotation and orientation of fluorescent molecules in plane polarized light, FP uses the rotation time and fluorescence lifetime for the molecules to determine the amount of filtered light transmitted to a detector (26). These characteristics can be exploited for immunoassay because the binding of a fluorescent (labeled) molecule to an antibody leads to the ordering of the complex with a concurrent and proportional increase in the transmittance of light through the filters. For example, the binding of two antibody molecules to a fluorescein-labeled protein A molecule can result in a complex having a molecular weight approximately eight times that of the original fluorescein-protein A complex (Fig. 7).

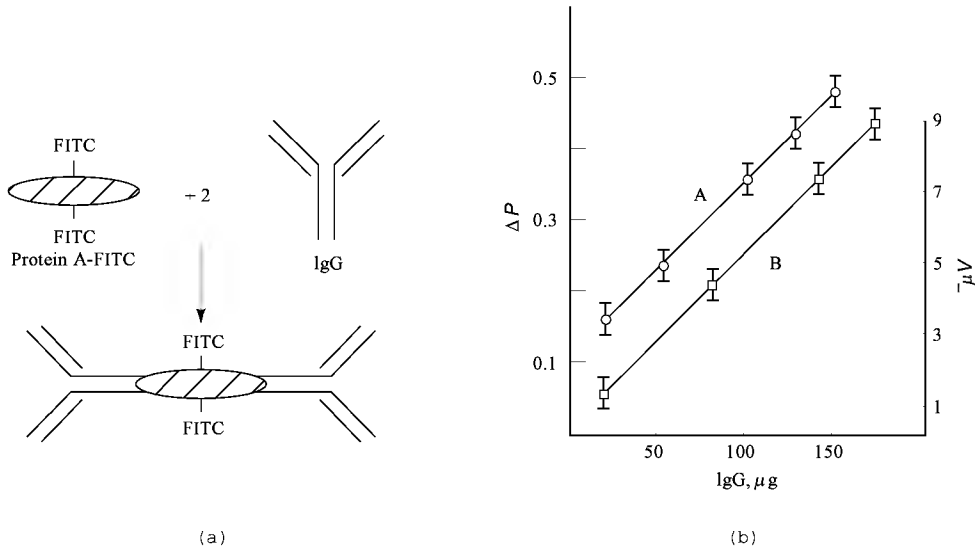


Fig. 7. Fluorescence polarization (FP). (a) The formation of the large FITC–protein A–IgG complex which leads to a net increase in plane polarized light transmitted from the solution. Molecular weights of the protein A-FITC, IgG, and complex are ca 43,000, 150,000, and 343,000, respectively. (b) Detection of IgG by fluorescence polarization immunoassay using A, a laboratory fluorimeter where (○) represents ΔP change in polarization, and B, a portable detection unit where (□) is μV change in voltage (27). The field detector proved to be more sensitive than the fluorimeter.

Immunoglobulin G (IgG) can be detected in less than one minute by FP using either a commercial fluorimeter or a portable field detector using a one-tube, homogenous format (Fig. 7b). Only one reagent needs to be prepared for this assay, αIgG -fluorescein. Fluorescein conjugates with antibodies are easily prepared by reaction of the antibody with fluorescein isothiocyanate at pH 9 for 3 h at room temperature followed by gpc purification. The resulting conjugate reacts directly with the sample containing IgG. Using this assay, ppm to ppb (μg to ng mL) amounts of IgG could be detected directly in blood serum in less than one minute. This same approach has been used for the direct measurement of viruses in less than one minute (19).

FP is also the basis for a very successful line of automated immunoanalyzers for the clinical measurement of application drugs and other analytes in a competitive mode (28). For example, to measure theophylline, drug antibody, theophylline–fluorescein conjugate, and the theophylline sample are combined. The labeled and unlabeled drug then compete for antibody. The amount of theophylline in the sample is then determined by the amount of polarized light transmitted by the solution. The higher the amount of drug in the sample, the less drug-fluorescent conjugate is bound to the antibody, and the less light is transmitted to the photometer-based detection system. As in all competitive binding assays, the amount of drug in a sample is inversely proportional to signal output, in this case transmitted light.

Other examples of FIA techniques include a fluorescence protection immunoassay, where a competition reaction is used between antibody, antigen–fluorophore, antibody to the fluorophore (antifluor), and free antigen such that the free antigen causes antifluor to bind to the antigen–fluor and fluorescent quenching occurs (29); and a homogenous FIA using a nonfluorescent drug-fluor conjugate as the substrate for an enzyme where free drug prevents antibody binding to the drug-fluor resulting in enzymatic degradation of the drug-fluor complex and fluorescence (30).

Chemiluminescent Immunoassay. Chemiluminescence is the emission of visible light resulting from a chemical reaction. The majority of such reactions are oxidations, using oxygen or peroxides, and among the first chemicals studied for chemiluminescence were luminol (5-amino-2,3-dihydro-1,4-phthalazinedione [521-31-3]) and its derivatives (see LUMINESCENT MATERIALS, CHEMILUMINESCENCE). Luminol or isoluminol can be directly linked to antibodies and used in a system with peroxidase to detect specific antigens. One of the first applications of this approach was for the detection of biotin (31).

In the most common method for chemiluminescent immunoassay (CLIA), after the immunological reaction and any necessary separation steps, the labeled compounds or complexes react with an oxidizer, eg, hydrogen peroxide, and an enzyme, eg, peroxidase, or a chelating agent such as hemin or metal

ions to produce light. This light is then detected and quantified using a photon or liquid scintillation counter (see [PHOTODETECTORS](#)).

A number of companies have developed and commercialized CLIA immunoassay systems. One system utilizes a competitive, heterogenous CLIA containing acridinium ester-labeled antigens or antibodies and micrometer-sized paramagnetic particles coated with antigens or antibodies (32). For an assay to measure an analyte such as ferritin [9007-73-2] or human chorionic gonadotrophin [9002-61-3] (hCG), the analyte in the sample competes with acridinium-labeled analyte for binding to the antibody on the magnetic beads. The beads are then magnetically separated, from the unbound antigens, and a chemiluminescent reaction occurs with the addition of hydrogen peroxide and sodium hydroxide. Photon emissions at 430 nm are then measured and compared to a standard curve to determine the sample concentration of unknown analyte.

Another commercial CLIA system uses antibody-coated beads to capture either unknown analyte or alkaline phosphatase-labeled analyte using a competitive heterogenous assay (33). Excess analyte and labeled analyte are then removed from the tube through a series of centrifugal washings. The amount of sample analyte is then determined by adding substrate (dioxetane) and measuring the emitted light as the substrate is hydrolyzed by the alkaline phosphatase.

Bioluminescence can also be used as the basis for immunoassay. For example, bacterial luciferase has been used in a co-immobilized system to detect and quantify progesterone using a competitive immunoassay format (34), and other luciferase-based immunoassays have been used to quantify insulin, digoxin, biotin, and other clinically important analytes (35).

Examples of Other Immunoassay Methods. A number of other immunoassay technologies have been developed but are not widely commercialized. These include liposome-mediated immunoassays, in which liposomes labeled with a fluorophore are used as the basis for detection (36); metalloimmunoassays, in which antigens or antibodies are labeled with a metal ion and then quantified after the immunoreaction using atomic absorption spectrophotometry or fluorometry (37); and a variety of magnetic bead-based immunoassays in which the application of a magnetic field aids separation of antigen–antibody complexes from free antigen during the assays (38). These are only a few of the many variations possible on the basic antigen–antibody binding reaction.

Comparison of Methodologies.

Heterogenous and Homogenous Immunoassays. The DAS EIA is an example of a heterogenous immunoassay ie, a multistep assay requiring the sequential addition of reagents with washing steps between reagent additions. In a typical protocol for the sandwich assay (see Fig. 1) and using an enzyme-based indicator, the sample containing the analyte is added to the immobilized antibody and permitted to react between 15 minutes to several hours, depending on the affinity of the antibody for the analyte. Repeated washings rinse away excess sample, and the second antibody linked to an enzyme is added. After another incubation (lasting 15 min–2 h) excess (nonbound) antibody–enzyme is washed away. A substrate is then added to the resulting two-antibody–analyte–enzyme complex, the action of the complexed enzyme on the substrate causes a color change, and that color is then quantified with a spectrophotometer. This whole process takes from 2 to 6 hours to complete. In competitive immunoassay (see Fig. 2) a similar series of reagent additions and washing steps would be required.

Most immunoassay kits and many commercial immunoassay analyzers are based on heterogenous EIA or FIA. These include an immunoassay system that uses FIA linked to radial partition chromatography of the antibody–antigen complex (39); a system that uses antibody-coated tubes for enzyme immunoassay of a variety of hormones and drugs (40); and a system that uses either a sandwich or competitive FIA format to measure a variety of analytes (41).

During the 1980s, a number of homogenous immunoassays were developed and commercialized. Homogenous immunoassays occur in one vessel, requiring no separation of components prior to quantification. Examples of homogenous immunoassays include a series of kits for stand-alone and automated immunoassays based on an enzyme activated assay using glucose-6-phosphate [56-73-5] (42), two very successful automated immunoanalyzers based on fluorescence polarization (FP) immunoassays and having a total menu of assays for over 100 analytes (43), and the FP technology developed and described earlier (19,27).

The advantages of homogenous immunoassays are simple formats and rapid data output producing user-friendly and cost-effective products. Technical challenges to consider, however, are the necessity to remove or minimize background interference from the reagents and nonspecific binding reactions.

Monoclonal vs Polyclonal Antibodies. A continuing question facing the developer of an immunoassay is whether to use monoclonal (MAb) or polyclonal (PAb) antibodies in the assay. Polyclonal antibodies are the natural mixture of antibodies resulting from the immune response to an antigen. A family of antibodies results, each binding specifically to a different antigenic determinant (or part of a determinant) on the same antigen. Subsets of polyclonal antibodies can also exist in which all the antibodies are specific for one antigenic site (epitope), but having varying avidities for that site.

Whereas such diversity in the immune response may have evolved as a protective means to the host animal, PABs present problems to the immunoassay developer looking for high antibody specificity and low total protein. For example, in most crude antisera, over 90% of the antibodies present have no or very low avidity for the antigen. This means that extensive purification must be used to isolate the specific antibodies. Routine purification measures cannot, however, separate the resulting specific antibodies further according to the specific determinant they bind to on the antigen. As a result, it is difficult to use the antibodies from the same preparation in an assay requiring two antibodies owing to differences in avidities between the antibodies and cross-reactivity problems.

In 1975, the first successful production of MABs was reported (44). By fusing normal antibody-producing cells with a B-cell tumor (myeloma), hybridoma cell lines resulted which produced antibodies having a specificity to only one determinant on an antigen; ie, all the antibodies produced from the cell line are identical. These studies resulted in a standard approach to MAB production. In this approach, the hybridoma cells are produced in large quantities in culture and screened to select specific clones producing the desired MAB using an appropriate assay. The selected clones are then expanded in culture (or in animals), the cells are collected, and the MABs are extracted and purified.

The singularity of MABs and the ease of mass production appeared to be the answer to rapid development of highly specific immunoassays. Companies were formed to produce MABs and incorporate them into assays. In fact, such assays have been developed and have proved very successful for infectious diseases, hormones, and other clinical analytes.

Whereas MABs appear to be the choice for use in immunoassays, a majority of immunoassay developers and suppliers use polyclonal antibodies. The primary reason for this choice lies in the investment of time and costs required to fuse, clone, and screen thousands of hybridomas to discover those producing MABs having the high avidity required for an assay. In most cases, PAB preparations with avidities equal to or better than the great majority of MABs produced from a fusion can be produced and purified in less time with far less investment. In addition, MAB-producing hybridoma cells can be extremely unstable, losing antibody production capabilities or simply dying out in a few passages (generations). Whereas PAB sources, such as animals and cell culture, are also susceptible to loss or change, these are more easily replicated.

The question of whether to use MABs or PABs in an assay is a matter of assay requirements (specificity and sensitivity) and economics and cannot be answered on technical merit alone.

Immunoassay–DNA Probe Hybrid Assays

Nucleic acid (deoxyribonucleic acid (DNA) and ribonucleic acid (RNA)) probes utilize labeled, ie, radioactive, enzymatic, or fluorescent, fragments of DNA or RNA (the probe) to detect complimentary DNA or RNA sequences in a sample. Because the probe is tailored for one specific nucleic acid, these assays are highly specific and very sensitive (45).

As the result of high specificity and sensitivity, nucleic acid probes are in direct competition with immunoassay for the analytes of some types of clinical analytes, such as infectious disease testing. Assays are being developed, however, that combine both probe and immunoassay technology. In such hybrid probe–immunoassays, the immunoassay portion detects and amplifies the specific binding of the probe to a nucleic acid. Either the probe per se or probe labeled with a specific compound is detected by the antibody, which in turn is labeled with an enzyme or fluorophore that serves as the basis for detection.

As of this writing, hybrid probe immunoassays are primarily in the laboratory development stage. Assays using the technology have been developed

for a number of bacteria, viruses, and human chromosome segments (46). For example, a method has been developed for the detection of hepatitis B virus DNA in serum using a sulfonated DNA probe. The DNA probe is then detected using (mouse) antibody to sulfonated cytosine and anti-mouse IgG conjugated to alkaline phosphatase. The resulting complex (DNA-sulfonated probe- α , sulfonated probe- α , α -sulfonated probe-alkaline phosphatase) is then incubated with an alkaline phosphate substrate to produce color (47).

Hybrid probe-immunoassays are expected to find a specific niche in clinical analysis, especially as a means to adapt probe assays to existing immunoanalyzers which are locked into a specific enzyme or fluorescence detection technology. Commercialization of the first of these assays is expected by the year 2000.

Immuno(bio)sensors

Immunoassay technology is also being applied in the development of antibody-based biosensors (qv), or immunosensors (27). A biosensor is an electronic detection device containing a biological molecule, such as an antibody, enzyme, or receptor, as its basic detection element. The biological molecule is immobilized onto a transducer that detects the immobilized biomolecule analyte interaction and sends a signal to the electronic portion of the biosensor for amplification and report output (Fig. 8). The ideal biosensor employs a homogenous (one-step, no-prep) format, real-time detection (results in less than one minute) in a cost-effective, portable, user-friendly design.

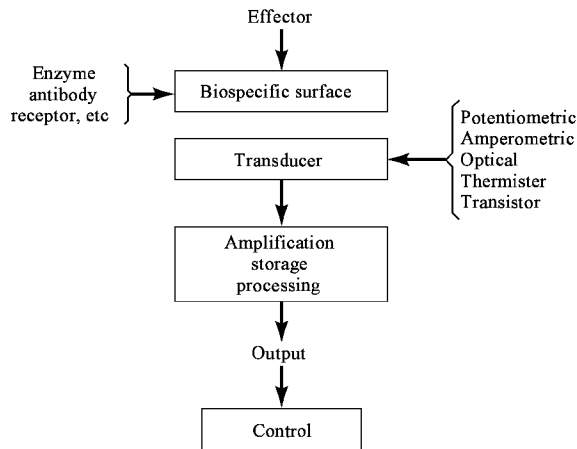


Fig. 8. Basic components of a biosensor. In the case of an immunosensor, the antibody (or antigen) would be immobilized onto the transducer.

Immunosensors have been designed which use both direct and indirect immunoassay technology to detect specific analytes within a minute or less in a variety of matrices (see Fig. 9). Indirect immunosensors may employ EIA, FIA, or CLIA principles whereby enzyme-, fluorophore- or chemiluminescent-labeled analyte competes with the target (nonlabeled) analyte for binding sites on the immobilized antibody. Unbound (free) labeled analyte is then quantitated using an electrochemical, optical, or electromechanical transducer and compared to the amount of target analyte in the sample.

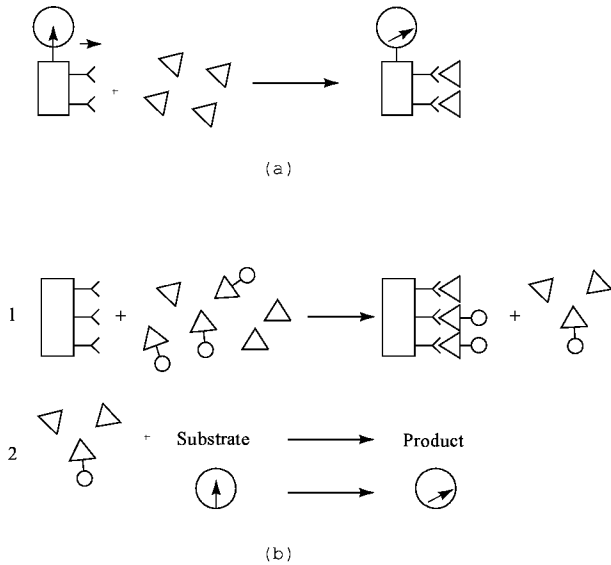


Fig. 9. Immunosensor approaches where Δ is the analyte, Δ is the labeled analyte, and Y is the antibody. (a) Direct immunosensors where the actual antigen-antibody interaction is measured; (b) indirect immunosensors 1 and 2 which utilize formats similar to competitive and displacement immunoassays.

Direct immunosensors measure the actual interaction between the immobilized antibody or antigen and the target analyte, which may be an antigen or antibody, respectively. Measurements may be based on the perturbation of an electrical field by the antibody-antigen binding event; changes in light scattering fluorescence or chemiluminescence on an optical fiber; or changes in the weight of an antibody-antigen complex as compared to the weight of antibody alone on a piezoelectric crystal.

A number of immunosensors have been developed that use an interdigitated electrode, capacitance transducer to measure direct antigen-antibody interactions (Fig. 10a) (48). The antibody (or antigen) is immobilized in a protein-based film covering one side of the chip, whereas a film containing no antibody (which serves as the background control) coats the other set of electrodes on the chip. The chip fits into a portable electronics module which supplies a low (ca 1 V) potential across the chip and which measures changes in impedance across the two sets of electrodes on the chip. After addition of sample, the impedance of the control film is subtracted from that of the antibody (or antigen) containing film to result in the measurement of specific binding. The output resulting is directly proportional to the amount of target analyte in the sample. Sample is added directly to the chip and the result read in approximately 20 seconds; thus this is a direct, homogenous immunosensor.

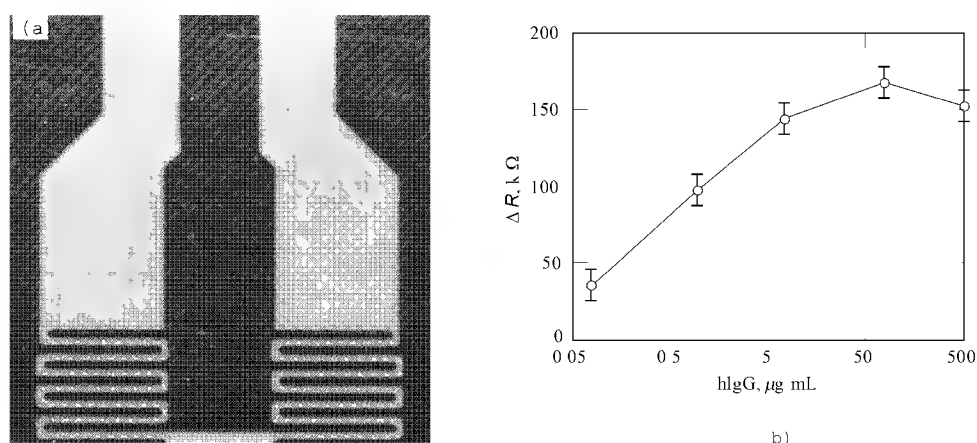


Fig. 10. (a) An interdigitated electrode biochip. One set of electrodes is coated with membrane containing antibody, the other set with membrane alone. The nonantibody containing membrane electrode serves as a control to compensate for nonspecific binding and background. (b) Response of biochip containing immobilized antibody to human IgG (αhIgG) to increasing amounts of hIgG in saline. The electronics unit reports specific antigen–antibody binding corrected for background and nonspecific binding as a change in impedance across the interdigitated electrodes (48).

Figure 10b illustrates the detection of human IgG (hIgG) by a biosensor. The bioactive film on the chip contained αhIgG entrapped with glutaraldehyde in a bovine serum albumin film. The chip was challenged with increasing concentrations of hIgG. Saline washings were used between challenges. As shown, the change in impedance is directly proportional to hIgG concentration in the ppb (ng mL) to ppm ($\mu\text{g mL}$) range.

Immunosensors promise to become principal players in chemical, diagnostic, and environmental analyses by the latter 1990s. Given the practical limits of immunosensors (low ppb or ng mL to mid-pptr or pg mL) and their portability, the primary application is expected to be as rapid screening devices in noncentralized clinical laboratories, in intensive care facilities, and as bedside monitors, in physicians' offices, and in environmental and industrial settings (49–52). Industrial applications for immunosensors will also include use as the basis for automated on-line or flow-injection analysis systems to analyze and control pharmaceutical, food, and chemical processing lines (53). Immunosensors are not expected to replace laboratory-based immunoassays, but to open up new applications for immunoassay-based technology.

BIBLIOGRAPHY

1. R. F. Taylor, *Worldwide Immunoassay Markets, Technologies and Applications, 1992–1998*, Decision Resources, Inc., Waltham Mass., 1994.
2. P. Ehdich and J. Morgenroth, *Dritte Mitteilung, Berl. Klin. Wochschr.* **37**, 453 (1900).
3. R. Kraus, *Wien. Klin. Wochenschr.* **10**, 736 (1897).
4. M. Heidelberger and F. E. Kendall, *J. Exp. Med.* **62**, 697 (1935).
5. P. Grabar and C. A. Williams, *Biochim. Biophys. Acta* **10**, 193 (1953).
6. A. S. Berson and R. S. Yallow, *Nature (Lond.)* **191**, 1392 (1961).
7. S. Avrameas, *Immunochemistry* **6**, 43 (1969).
8. J. E. Butler, in E. T. Maggio, ed., *Enzyme Immunoassay*, CRC Press, Boca Raton, Fla., 1981, pp. 5–52.
9. S. F. de St. Groth, in I. Lefkovits and B. Pernis, eds., *Immunological Methods*, Academic Press, Inc., New York, 1979, pp. 1–42.
10. B. F. Erdanger, *Meth. Enzymol.* **70**, 85 (1980).
11. D. S. Kabakoff, in Ref. 8, pp. 71–104.
12. R. F. Taylor, in R. F. Taylor, ed., *Protein Immobilization*, Marcel Dekker, Inc., New York, 1991, pp. 139–160.
13. N. M. Green, *Meth. Enzymol.* **18**, 418 (1970).
14. O. Ouchterlony, in *Handbook of Experimental Immunology*, Blackwell, Inc., London, 1967, pp. 655–724.
15. J. M. Singer and C. M. Plotz, *Am. J. Med.* **21**, 888 (1956).
16. L. B. Bangs, *Am. Clin. Lab.*, 5 (Aug. 1988).
17. J. E. Coligan, T. J. Kindt, and R. M. Krause, *Immunochem.* **15**, 755 (1978).
18. Pacific Biotech, Inc., *CARDS-OS Strep A Direct Antigen Test*, Bulletin IN7740-B, 1992.
19. R. F. Taylor, *Advanced Immunological Techniques for Biological Detection*, U.S. Army CRDC Document #CRDC-CR-84-113, Washington, D.C., 1984.
20. A. Johansson, C. J. Stanley, and C. H. Self, *Clin. Chim. Acta* **148**, 119 (1985).
21. D. L. Morris and co-workers, *Anal. Chem.* **53**, 658 (1981).
22. C. Cox and J. Buret, *Anal. Chem.* **54**, 1862 (1982).
23. A. M. Coons and M. M. Kaplan, *J. Exp. Med.* **91**, 1 (1950).
24. H. M. Shapiro, *Cytometry* **3**, 227 (1983).
25. I. Hemmila, *Clin. Chem.* **31**, 359 (1985).
26. W. B. Dandliker and co-workers, *Immunochem.* **10**, 219 (1973).
27. Ref. 12, pp. 263–303.
28. M. E. Jolley and co-workers, *Clin. Chem.* **27**, 1190 (1981).
29. R. F. Zuk, G. L. Rowley, and E. F. Ullman, *Clin. Chem.* **25**, 1554 (1979).
30. J. F. Burd and co-workers, *Clin. Chem.* **23**, 1402 (1977).
31. H. R. Schroeder and co-workers, *Anal. Chem.* **48**, 1933 (1976).
32. R. F. Dudley, *J. Clin. Immunoassay* **14**, 77 (1991).
33. A. L. Babson, *J. Clin. Immunoassay* **14**, 83 (1991).
34. B. Terouanne and co-workers, *Analyt. Biochem.* **154**, 118 (1986).
35. F. Gorus and E. Schram, *Clin. Chem.* **25**, 512 (1979).
36. J. P. O'Connell and co-workers, *Clin. Chem.* **31**, 1424 (1985).
37. M. S. Cais and co-workers, *Nature* **270**, 534 (1977).
38. O. Olsvik and co-workers, *Clin. Microbiol. Rev.* **7**, 43 (1994).
39. D. S. Plaut and W. N. McLellan, *J. Clin. Immunoassay* **14**, 120 (1991).
40. R. Jefferson, *J. Clin. Immunoassay* **14**, 89 (1991).
41. E. Soini and H. Kojola, *Clin. Chem.* **29**, 65 (1983).
42. N. Bellet and B. S. Wagman, *Am. Clin. Prod. Rev.*, 10–15 (June 1986).
43. C. H. Keller, K. L. Fitzgerald, and A. Barnes, *J. Clin. Immunoassay* **14**, 115 (1991).

44. G. Kohler and C. Milstein, *Nature* **256**, 495 (1975).
45. C. D. Sauls and T. Caskey, *Clin. Chem.* **31**, 804 (1985).
46. J. L. Guesdon, *J. Immunol. Met.* **150**, 33 (1992).
47. A. Akar, B. Bournique, and R. Scholler, *Clin. Chem.* **38**, 1352 (1992).
48. R. F. Taylor, I. G. Marenchic, and R. H. Spencer, *Anal. Chim. Acta* **249**, 67 (1991).
49. L. Campanella and M. Tomassetti, *Select. Electrode Rev.* **11**, 69 (1989).
50. D. Monroe, *Crit. Rev. Clin. Lab. Sci.* **27**, 109 (1989).
51. G. Wagner and R. D. Schmid, *Food Biotech.* **4**, 215 (1990).
52. M. Behizad and co-workers, *Process Biochem.*, 126–132 (Aug. 1989).
53. M. T. Reilly and M. Charles, *Bioprocess Technol.* **6**, 243 (1990).

General References

G. E. Abraham, ed., *Handbook of Radioimmunoassay*, Marcel Dekker, Inc., New York, 1977.
J. Clausen, *Immunochemical Techniques for the Identification and Estimation of Macromolecules*, 3rd ed., Elsevier, Amsterdam, the Netherlands, 1988.
A. Kawamura, Jr., ed., *Fluorescent Antibody Techniques and Their Applications*, University of Tokyo Press, Baltimore, 1977.
L. J. Kricka, ed., *Analytical Applications of Bioluminescence and Chemiluminescence*, Academic Press, Inc., New York, 1984.
E. T. Maggio, ed., *Enzyme Immunoassay*, CRC Press, Boca Raton, Fla., 1981.
R. F. Taylor, ed., *Protein Immobilization: Fundamentals and Applications*, Marcel Dekker, Inc., New York, 1991.

Richard F. Taylor
Arthur D. Little, Inc.



IMMUNOSUPPRESSANTS.

See ANALGESICS, ANTIPYRETICS, AND ANTIINFLAMMATORY AGENTS; CHEMOTHERAPEUTICS, ANTICANCER; IMMUNOTHERAPEUTIC AGENTS.

IMMUNOTHERAPEUTIC AGENTS

Advances in immunology during the last part of the twentieth century have continued at a rapid rate and cytokines and immune cells having specific markers continue to be defined. A number of natural and synthetic immunotherapeutic agents have been discovered that can modulate components of the normal or aberrant immune system, through stimulation or suppression. However, most of these substances also have inherent adverse side effects.

The immune system (1) is the primary mechanism of defense against invasive disease for human beings. Microorganisms can penetrate into the innermost parts of the body and once there, if not combatted, can cause disease. A competent immune system can recognize the molecular components of these organisms or substances released by them. These highly specific molecules, called antigens (Ag), can provoke immune responses. In addition to microorganisms, chemicals such as immunoglobulins (Igs), nucleic acids (qv), and other biopolymers (qv); eg, blood cells or plasma proteins, introduced through transfusions (see FRACTIONATION, BLOOD); food; and blood substitutes may be important sources of antigens. Moreover, internal tissue injuries induced physically or chemically can release antigenic material. Cell turnover, which may be accelerated under various conditions, can also lead to an increased antigen load. Many of the biopolymers released during cell death may be denatured or partially metabolized to become more antigenic. As a body's anabolic and catabolic processes fluctuate, the availability of these antigens may rise and fall with the state of health and disease. During aging, active stages of diabetes, starvation, or even a crash-diet program, deterioration of tissue components may expose the immune system to many antigens. The ability of the immune system to respond determines the difference between normalcy, ie, health, and pathology, ie, disease.

The immune system is composed of two principal components, known as limbs: the humoral immune system, primarily the domain of the B-lymphocytes; and the cell-mediated immune system, primarily the domain of the T-lymphocytes. In a normal, healthy state, regulatory mechanisms affect control so that these two limbs function in proper balance. The humoral immune system produces antibodies that react specifically with antigens; the cell-mediated immune system mobilizes the phagocytic leucocytes to ingest and destroy invading organisms or molecular material that contain or release the antigens. In the healthy state, the two limbs are regulated by complex feedback mechanisms employing mediators called cytokines and by cell-to-cell cooperation. Sudden exposure to antigens can elicit responses by the immune system such as clonal proliferation, an expansion process by which a few antigen-responsive cells can generate a large number of specific immune competent cells.

Whereas the rate and extent of an immune response may depend on the specific cells available at the moment an antigenic challenge occurs, the capacity to mount an immune response is largely determined genetically. For example, T-cell recognition of antigens, the ability of cells to present antigen, and the potential of the B-cell to produce antibodies are regulated by immune response genes. The ability to control immune responses, and thereby avoid excessive production of antibodies to self-antigens, is critical to prevention of autoimmunity.

Cytokines and antagonists (2–4), intercellular proteins produced by immune cells, play an important role in the regulation of immune responses. Cytokines are present in a variety of tissues under normal conditions. Through insufficient or excessive production, these macromolecules can mediate chronic inflammatory diseases. An inability to respond to cytokines, eg, interleukin 1 (IL-1) or interleukin 2 (IL-2), may lead to an immunosuppressive state, whereas over-production can result in severe shock, autoimmune disease, or immunopathological conditions, such as leukemia and rheumatoid arthritis (RA). Specific communications between immune cells are constantly modulated by naturally occurring inhibitors.

The number of known cytokines, as well as the diversity of biological functions, have led to a very complex and often confusing picture of the immunologic and nonimmunologic processes involved. The role of cytokines in local or systemic homeostatic mechanisms related to physiological functions may be utilized therapeutically for treatment of cancer and a variety of other diseases (2). Pharmaceutical research and development efforts surrounding IL-1 are typical examples of the cytokine inhibition approach to chronic inflammation research (2).

T-Lymphocytes (4,5) and other cellular components of the immune system also have equally wide implications in regulation of the normal immune system. The T-lymphocytes play a central role in the body's response to harmful antigens and tumor–host interaction (4). Responses involve antigens derived from viruses, bacteria, parasites, and tumors. T-cells also participate in the immune surveillance response, where self-antigens are recognized, but usually sequestered within the cell and, when exposed, become markers of cellular damage.

Because T-lymphocytes are capable of recognizing and destroying pathologically altered self-tissues, it may be possible to utilize T-lymphocytes for treatment of malignancy (2) or chronic and debilitating autoimmune diseases such as rheumatoid arthritis, thyroiditis, and neuromuscular diseases (5). Clinical studies for the use of T-cell lines for vaccination and treatment of autoimmune diseases, contemplated since the early 1980s, were initiated in 1990 at the Brigham and Women's Hospital (Boston, Massachusetts) (see VACCINE TECHNOLOGY).

Immunodeficiencies

Primary immunodeficiencies are uncommon, and may occur in 1 in 10,000 individuals (6). Many primary immunodeficiencies are hereditary and congenital, and first appear in infants and children. Primary immunodeficiencies are classified into four main groups (7) relating to the lymphocytes (B-cells, T-cells, or both), phagocytes, or the complement cascade (8). Primary deficiency diseases result from B-cell defects in 50% of cases, from T-cell defects in ca 10%, and from combined B- and T-cell defects in ca 20%. Phagocytic disorders account for 18% and complement defects occur in 2% of all cases.

In severe combined immunodeficiency disease (SCID), B- and T-cell immunity are either absent or depressed. Children having SCID are susceptible to infection from almost any microorganism and often die within their first year of life, unless placed in a sterile environment. Individuals with the same primary immunodeficiency may express different signs and symptoms. In general, chronic or recurrent infections and unusual or selective infections suggest primary immunodeficiency. Quantitation of serum antibody levels and isotyping of the serum immunoglobulin (IgG) levels are required to confirm primary deficiency disease. A deficiency in one subclass can result in signs and symptoms of immunodeficiency. Treatment consists of a maintenance dose of immune globulin intravenous (IGiv), usually 150–300 mg/kg every 3–4 wk. Higher frequency of dosing may be needed for some patients to maintain serum antibody levels at a concentration of 200 mg/dL. For acute infections, 500 mg/kg may be infused along with antibiotics (qv).

Secondary immunodeficiencies (9) are much more common than primary ones and frequently occur as a result of immaturity of the immune system in premature infants, immunosuppressive therapy, or surgery and trauma. Illnesses, particularly when prolonged and serious, have been associated with secondary immunodeficiencies, some of which may be reversible. Acquired immune deficiency syndrome (AIDS) (10–12) may be considered a secondary immunodeficiency disease caused by the human immunodeficiency viruses HIV-1 or HIV-2. Hitherto unknown, the disease began to spread in the United States during the latter part of the 1970s. The agent responsible for this infection has been isolated and identified as a retrovirus.

The pathogenesis of AIDS (10,12,13) following HIV infection may be separated into primary and secondary effects. The primary effects are (1) quantitative and qualitative decreases in infected cells, ie, the CD4⁺ T-lymphocytes; (2) impaired cellular immunity; (3) impaired immune surveillance; and

(4) direct pathology in the central nervous system (CNS), gut, and elsewhere. Secondly there are devastating effects of opportunistic infections and appearance of virally mediated tumors.

Gradual diminution of CD4⁺ T-lymphocytes from the peripheral blood is the most consistent feature observed in HIV infection. Because the majority of CD4⁺ cells are T-helper lymphocytes, removal leads to deficiency of cellular immunity, which depends on T-helper cells to initiate cytotoxic T-cell killing of virus-infected cells of cancer. The loss of immune surveillance leads to the appearance of virally induced tumors from unopposed clonal expansion of virally transformed cells. Furthermore, depletion of cellular immunity leads to exaggerated viral, fungal, and protozoal infections.

Pneumocystis carini pneumonia (PCP), the most common of the opportunistic infections, occurs in more than 80% of AIDS patients (13). Toxoplasmosis, a protozoan infection of the central nervous system, is activated in AIDS patients when the CD4⁺ count drops and severe impairment of cell-mediated immunity occurs. Typically, patients have a mass lesion(s) in the brain. These mass lesions usually respond well to therapy and can disappear completely. Fungal infections, such as *Cryptococcal meningitis*, are extremely common in AIDS patients, and *Histoplasma capsulatum* appears when cell-mediated immunity has been destroyed by the HIV virus, leading to widespread infection of the lungs, liver, spleen, lymph nodes, and bone marrow. AIDS patients are particularly susceptible to bacteremia caused by nontyphoidal strains of *Salmonella*. Bacteremia may be cleared by using antibiotic therapy.

Immunotherapy for Various Disease States

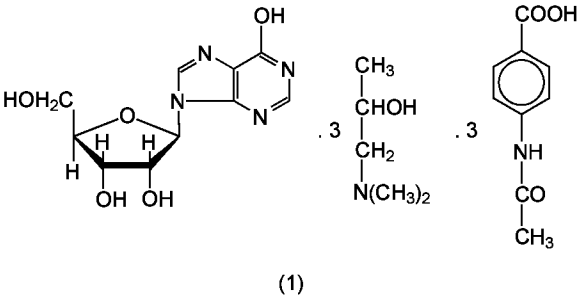
Immunodeficiencies. Whereas elimination and neutralization of the antigen by avoidance and isolation is feasible for minor cases of allergies (14), patients with immunodeficiency are at greater health risk and require increased amount of care to maintain optimal health. Antibiotic (qv) therapy is an important modality for the treatment of immunodeficient patients and continuous prophylactic antibiotic treatment is often beneficial, especially when there is a potential for overwhelming infection, other forms of immune therapy have proven insufficient, or there is a high risk for a specific infection. The objective is the elimination of the organism or antigen as well as physiologically active substances that are not immunologically related. Antibiotics are lifesaving in treatment of infections, particularly in immunosuppressed or immunodeficient patients, and the selection and dosages used are critical.

In passive immunotherapy immune globulin (Ig) is an effective replacement in most forms of antibody deficiency (14). In the past, plasma was used instead of immune globulin, but plasma is rarely indicated in the 1990s because of the risk of disease, particularly AIDS, transmission. Because plasma contains many factors in addition to immunoglobulins (Igs), plasma is, however, of particular value in patients with protein-losing enteropathy, complement deficiencies, and refractory diarrhea.

Problems associated with active immunotherapy for a faltering immune system involve identifying the requirements for appropriate immunotherapeutic agents so that both efficacy and safety can be ensured. Immunological disturbances in patients can lead to pathogenesis of many diverse problems, such as senescence, primary and acquired immunodeficiency syndromes, acute and chronic infections, autoimmune diseases, and cancer. Immunoenhancing agents may be useful where a certain degree of immune capacity is present, but are of limited value in treating cellular or phagocytic immunodeficiencies.

Antiviral agents (qv) (15–17) are used in attempts to combat the devastating effect of HIV on the immune system. As of this writing there are three principal approaches to the treatment of AIDS: (1) use of anti-HIV agents to destroy the virus or control its growth; the National Cancer Institute (NCI) encourages submission of synthetic and characterized natural products for anti-HIV screening (18); (2) immunotherapy to restore impaired immune functions; and (3) treatment of specific opportunistic infections or tumors.

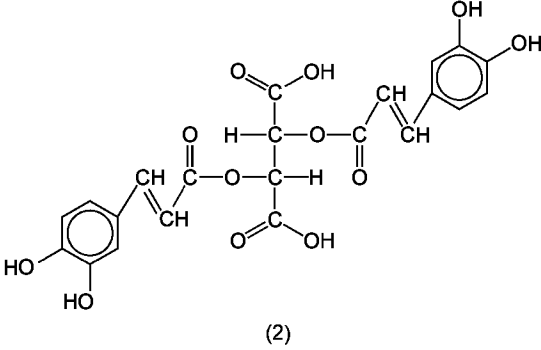
The majority of antiviral agents, whether purines, pyrimidinones, amantadines, or others, were designed primarily to interfere with viral replication. Purine (adenosine, guanosine, and inosine) and pyrimidine (cytidine and thymidine) nucleosides having the dideoxy configurations of the ribose moiety nucleosides, were shown in cell cultures to control the rapid proliferation of viruses (11,12). However, there are few virus-specific enzyme systems available as targets for appropriate chemotherapeutic intervention without concomitant adverse effects on host cellular processes. Usually antiviral chemotherapy results in a wide variety of side effects and a narrow separation between efficacy and toxicity. Some antiviral agents, eg, inosine pranohex [36703-88-5], C₅₂H₇₈N₁₀O₁₇ (1), are also immunomodulators.

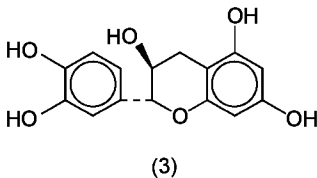


Plants and microorganisms produce unique and diverse chemical structures, some of which act as immunomodulators (18–28). Of specimens used in traditional medicine, approximately 450 plant species have shown antiviral activity out of 4000 plants screened (19). Several tannins (20) exhibit strong inhibition of tumor promotion experimentally. Pretreatment of mice with small amounts of tannins for several days strongly rejected transplanted tumors. This activity has been claimed to be effected through enhancement of host-mediated antitumor activity.

Whereas over 200 plant constituents are reported to have antiviral activity, as determined by *in vitro* methods, only 31 compounds have shown antiviral activity *in vivo* (19). Immunotherapeutic activity has not been determined.

A number of natural products exhibit immunological effects. Echinacea extracts contain a large number of constituents and the immunostimulatory activity of such extracts has been demonstrated (22). The biological activity of these extracts has warranted development of fermentation methods for large-scale production. Significant quantities of polysaccharide mixtures have been produced, leading to the isolation and availability of pure polysaccharides for biological testing. Lipophilic alkylamides and the polar caffeic acid [331-39-5] derivative, C₉H₈O₄, cichoric acid (2) (2,3-O-dicaffeoyltartaric acid), probably contribute to the immunostimulatory activity of the Echinacea extracts.

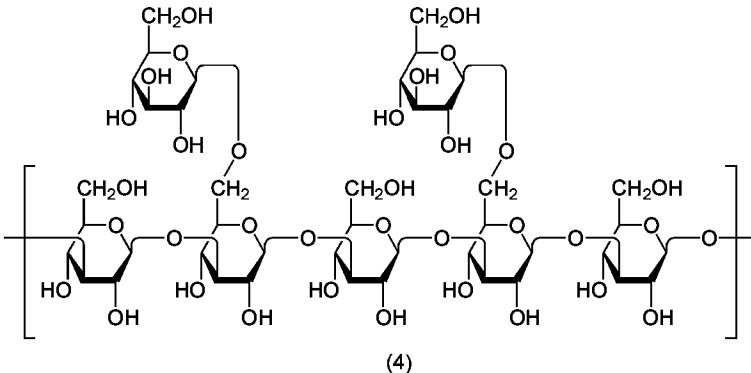




Cianidanol (Ci) or catechin [154-23-4], C₁₅H₁₄O (3), a flavonoid found primarily in higher woody plants (23), has been shown to have both some specific and nonspecific effects on the immune system (24). Cianidanol exerts an immunoenhancing effect on the function of various peripheral mononuclear blood cells (25). The T-cell activation by Ci at low doses stimulates the spontaneous and mitogen-induced proliferation and Ig secretion of human peripheral blood mononuclear cells. Ci is not able to stimulate B-cells directly, and T-cells are required for enhancement. The immunoenhancing properties of Ci were demonstrated (26) in a clinical study involving healthy volunteers and patients suffering from chronic liver diseases. After treatment for six months using 2- of Ci, the helper-to-suppressor T-lymphocyte ratio was significantly augmented. Some patients having high initial ratios showed a decrease with treatment, and T-suppressor cell concentrations were decreased at the same time. Significant clinical improvements were observed (27,28) in a four-month-long, double-blind, multicenter trial involving 338 patients with chronic active hepatitis. The clinical improvements may be related to the immunoenhancing properties of this agent.

Ling Zhi-8 (LZ-8) is an immunomodulatory protein (29,30) isolated from the mycelial extract of *Ganoderma lucidium*, that has been purified and shown to stimulate mouse spleen and human peripheral blood lymphocytes. LZ-8 is able to inhibit antibody production and prevent the development of autoimmune type I diabetes in NOD mice.

Polysaccharides (31–41) obtained from different sources have been shown to have immunostimulant and antitumor activity. One example is a glucan, isolated from yeast, a branched β-1,3-polyglucopyranose originally present in the yeast cell wall as a component of zymosan. It is known to be a broad-spectrum enhancer of host defense mechanisms (29–33). Immunopharmacological studies of this glucan demonstrated antitumor effects; prevention of carcinogenesis; increase in host resistance to bacterial, viral, fungal, and parasitic infections; and an increase in phagocytic and proliferative activity of the reticuloendothelial system. Lentinan [9051-97-21] (4), an adjuvant polysaccharide isolated from mushrooms and described as β-1,3-D-glucan having β-1,6-glucopyranoside branchings, has been shown to have immunomodulatory activity also.



Glucans have been shown to nonspecifically activate cells of the macrophage–monocyte series (34) and stimulate humoral (35) and cell-mediated immunity (36). Numerous studies have demonstrated that glucans profoundly enhance the production of colony-stimulatory factor (37) and the subsequent stimulation of diverse bone marrow progenitor cells (38,39). In addition, treatment using glucan inhibited the growth of a variety of experimentally induced syngeneic neoplasms and enhanced survival in four syngeneic murine tumor models (40). Other studies have shown that therapeutic application of glucan increased long-term survival, inhibited metastases, decreased primary tumor weight, and maintained hepatic parenchymal cell functional integrity in animals with syngeneic reticulum cell sarcoma M5076 (41). The tumoricidal activity of splenic and peritoneal macrophage cytolytic activity against M5076 was increased after three iv doses of glucan. Clinically, glucan appears to be an immunomodulator with hematopoietic and radioprotective capabilities.

Synthetic immunomodulators (42) that have been developed are listed in Table 1. Structures are given in Figure 1. These compounds have been shown to modulate the immune system. Some also have antitumor activity. Inosiplex [36703-88-5] (1), also known as inosine pranobex, has been used to treat immunodeficiencies caused by cancer, radiotherapy, surgery, burns, aging, and AIDS (43). Treatment with inosiplex, for one week to several months, reduced the incidence of complications, infections, and mortality, ie, the immune status of these patients was enhanced. Natural killer cell cytotoxicity was improved; T-lymphocyte count, mitogen-induced proliferation, E rosettes, and skin test reactivity also improved. A study of immunosuppressed male homosexuals, most of whom presented clinical signs and symptoms of prodromal AIDS, indicated that the patients' depressed immune parameters returned to normal. Treatment with inosiplex was found to have a profound and lasting effect on a number of immunological parameters.

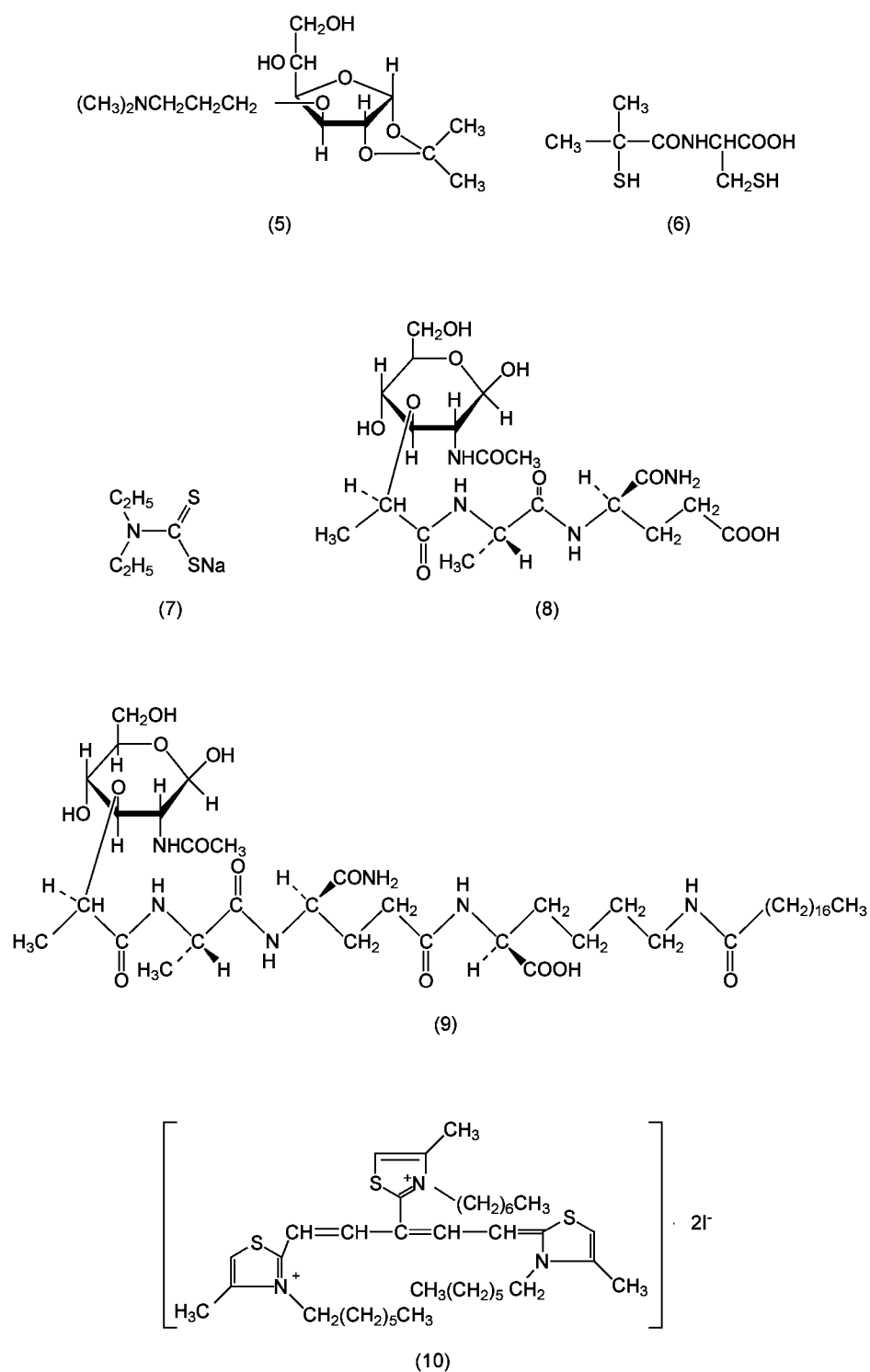


Fig. 1. Structures of synthetic immunomodulators. See Table 1.

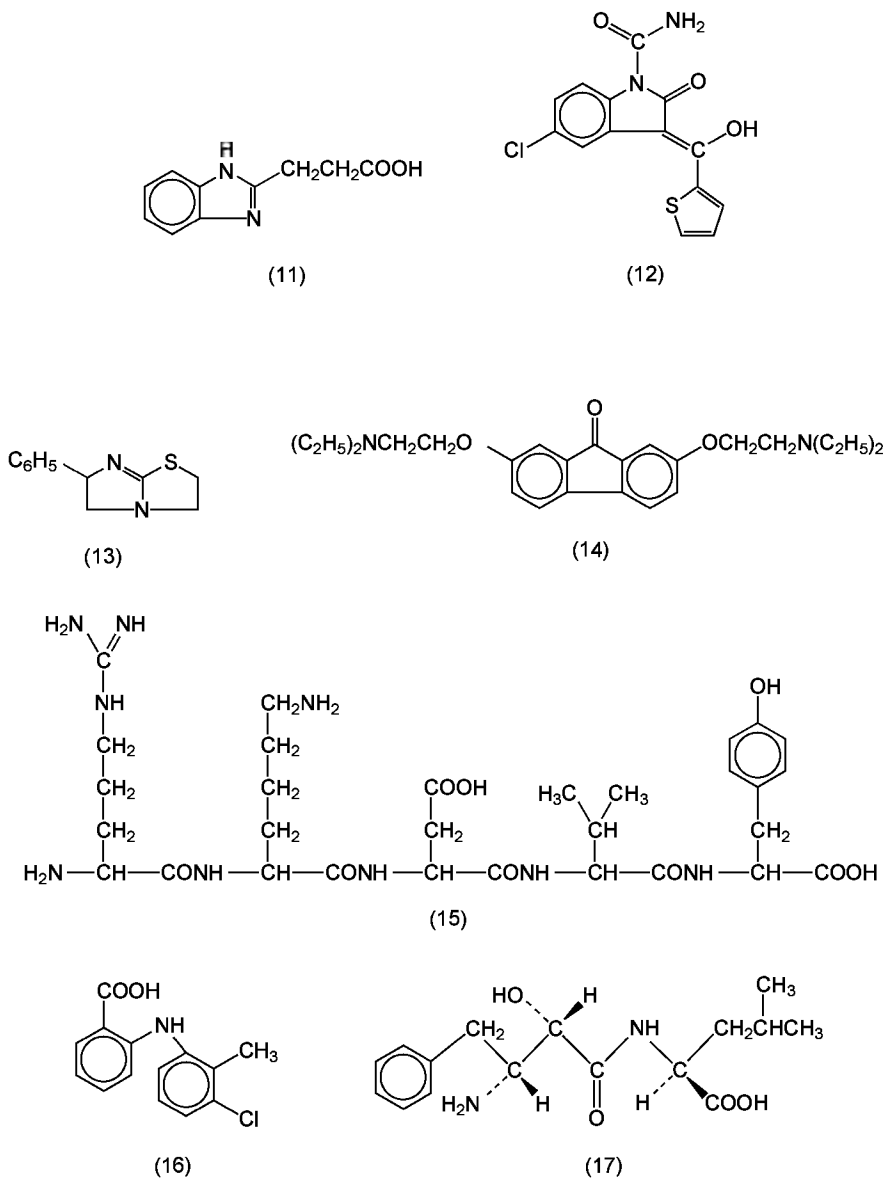


Table 1. Synthetic Immunomodulators^a

Compound	CAS Registry Number	Molecular formula	Structure number
amiprilose	[56824-20-5]	C ₁₄ H ₂₇ NO	(5)
bucillamine	[65002-17-7]	C ₇ H ₁₃ NO ₃ S ₂	(6)
ditiocarb sodium	[148-18-5]	C ₅ H ₁₁ NS ₂ ·Na ^b	(7)
inosine pranobex	[36703-88-5]		
muramyl dipeptide	[53678-77-6]	C ₁₉ H ₃₂ N ₄ O ₁₁	(8)
muroctasin	[78113-36-7]	C ₄₃ H ₇₈ N ₄ O ₁₃	(9)
platonin	[3571-88-8]	C ₃₈ H ₁₁ N ₃ S ₃ ·2I	(10)
procodazole	[23249-97-0]	C ₁₀ H ₁₀ N ₂ O ₂	(11)
tenidap	[120210-48-2]	C ₁₄ H ₉ ClN ₂ O ₃ S	(12)
tetramisole	[5036-02-2]	C ₁₁ H ₁₂ N ₂ S	(13)
tilorone	[27951-97-5]	C ₂₅ H ₃₄ N ₂ O ₃	(14)
thymopentin	[69558-55-0]	C ₃₀ H ₄₉ N ₉ O ₉	(15)
tolfenamic acid	[13710-19-5]	C ₁₄ H ₁₂ ClNO ₂	(16)
ubenimex	[58970-76-6]	C ₁ H ₂₄ N ₂ O ₄	(17)

^a See Figure 1.
^b C₁₀H₁₂N₄O₅·3C₉H₉NO₃·3C₅H₁₃NO.

The immunorestorative potential of inosiplex has been evaluated in several clinical conditions, including post-surgical trauma, cancer patients with concurrent viral infections, and cancer patients receiving radiotherapy or chemotherapy. For example, most (84%) of the surgery patients remained immunologically depressed, but 56% of the inosiplex-treated surgery patients had complete restoration of normal skin test reactivity (probability level < 0.0005). The use of inosiplex as an adjuvant to chemotherapy or radiotherapy appears to be valuable in the prophylaxis against opportunistic infections.

Imuthiol (7), ditiocarb sodium, was first studied in the early 1980s (44–49) when animal data suggested that Imuthiol is a virtually nontoxic immunotherapeutic agent. It is active on the T-cell lineage and is devoid of immunosuppressive effects. This compound induces T-cells to generate enhanced levels of cytotoxic activities, lymphoproliferative responses, and IL-2 production. Imuthiol increases natural killer cell (NK) activity without effects on circulating interferon levels. Through its effects on T-cells, B-cells are induced to secrete primary antibodies of the IgG class and monocyte macrophages to participate in delayed-type hypersensitivities, and to increase IL-1 production. Imuthiol can restore T-cell activities inhibited by cytotoxic agents. Imuthiol treatment in AKR mice, NZB mice, and virally induced diabetes or cancer has shown favorable results.

Generally, clinical studies using Imuthiol have confirmed the animal results. The compound was found to be safe when administered orally or intravenously (2–20 mg kg body weight) once a week for up to four years, without untoward side effects or incompatibility with other drugs. Imuthiol can restore abnormal T-helper:suppressor cell ratios, responses to T-cell mitogens, and delayed hypersensitivity. Immune restoration and beneficial effects were observed in chronic bronchitis, bronchiectasis, repeated upper respiratory infections, tuberculosis, post-surgery infections, infections in the elderly, and rheumatoid arthritis (49). Treatment with Imuthiol in AIDS patients (46,47) was followed by increased T-helper cell count and restoration of helper:suppressor cell ratios to normal. Three patients with lymphadenopathy associated syndrome were treated with Imuthiol with encouraging results. The use of Imuthiol as preventive therapy has been suggested for treatment of HIV seropositive subjects, particularly when they present a decrease in T-helper cell subset.

Rheumatoid Arthritis. Nonsteroidal antiinflammatory drugs (NSAIDs) are able to modulate many symptoms of chronic inflammation, but are unable to halt the underlying degenerative changes involved in rheumatoid arthritis (RA) (49–61), a common disease that affects $2-3 \times 10^6$ people in the United States alone (see ANALGESICS, ANTIPYRETICS, AND ANTIINFLAMMATORY AGENTS). In general, there is no evidence of immunological effects for the group of classical NSAIDs; eg, NSAIDs show no significant effect on the synthesis of IL-1, and use of NSAIDs in an autoimmune disease, such as RA, has been questioned. As of this writing, however, the NSAIDs continue to be the first-line drugs for the traditional treatment of chronic inflammation, including RA. When NSAIDs are not sufficiently effective or beneficial, a number of second-line agents, known as disease-modifying agents, are available.

Steroids, also used for RA treatment, generally influence the synthesis and response to IL-1. Glucocorticoids, eg, prednisone [53-03-2] (18), can affect virtually every aspect, phase, and cell type involved in immunologic and inflammatory reactions (56). Some rheumatologists are now using immunosuppressive drugs such as methotrexate [59-05-2] (19) in early stages of RA with significant success. Antimalarials, gold compounds, penicillamine, and sulfasalazine are all used as antirheumatics. Traditional antirheumatic drugs having immunological activity are listed in Table 2 and structures are given in Figure 2. Immunosuppressive drugs that are increasingly used for treatment of severe and active RA are given in Table 3. Structures are shown in Figures 3 and 4.

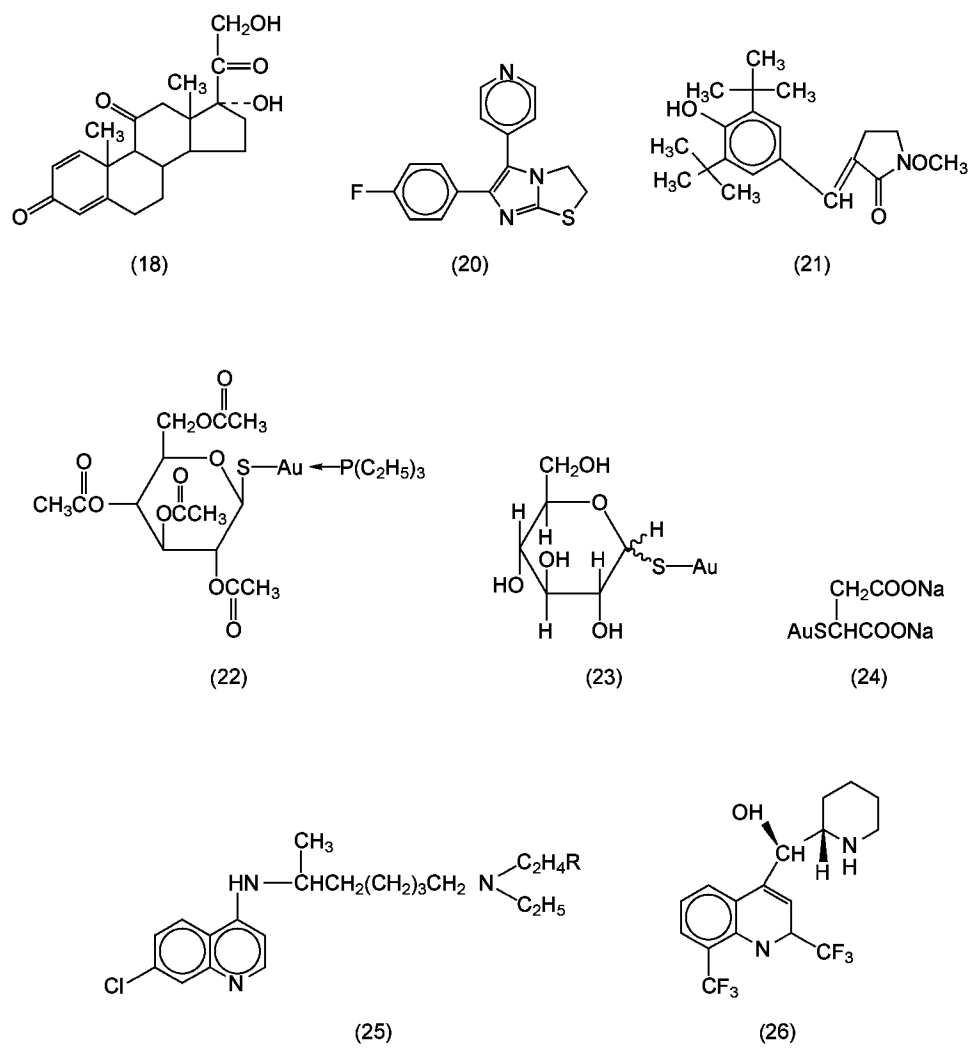


Fig. 2. Structures of traditional antirheumatic drugs having immunologic activity. See Table 2.

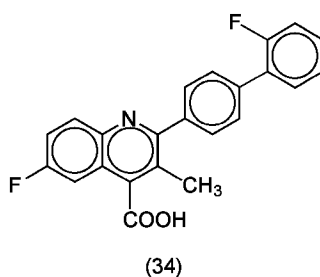
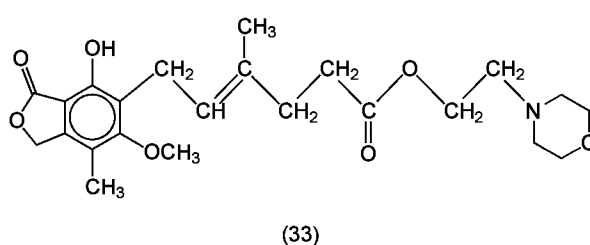
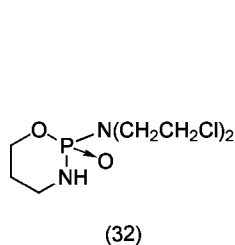
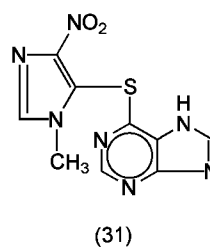
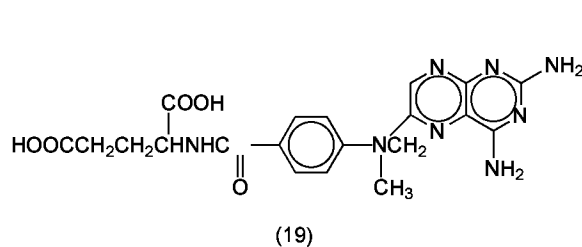
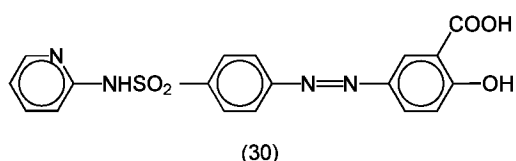
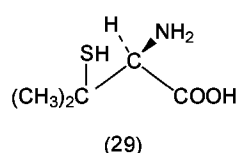
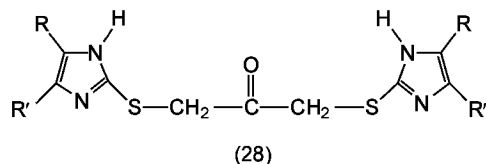
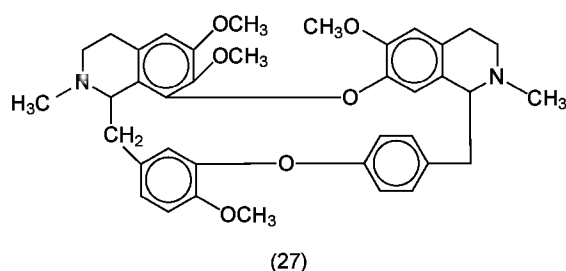


Fig. 3. Structures of synthetic immunosuppressive drugs. See Table 3.

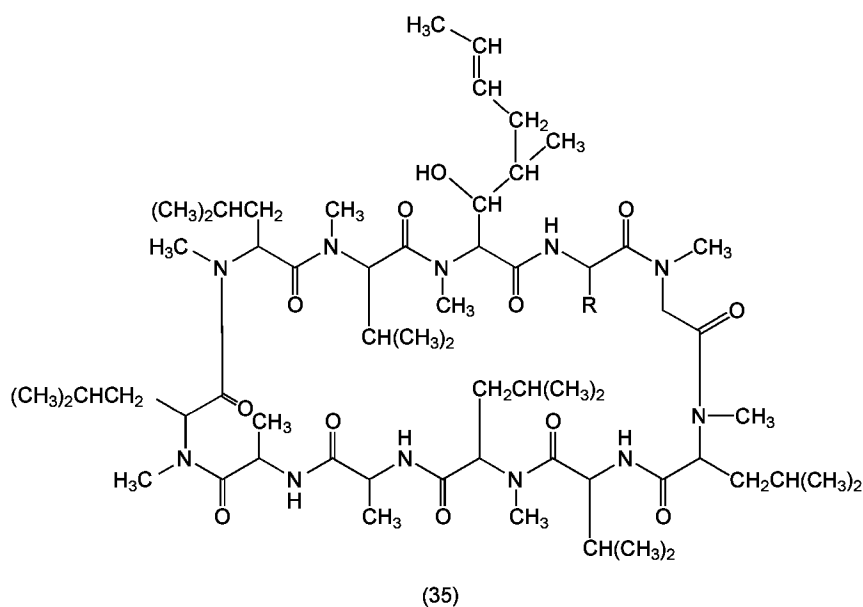


Fig. 4. Structures of natural immunosuppressive drugs. See Table 3.

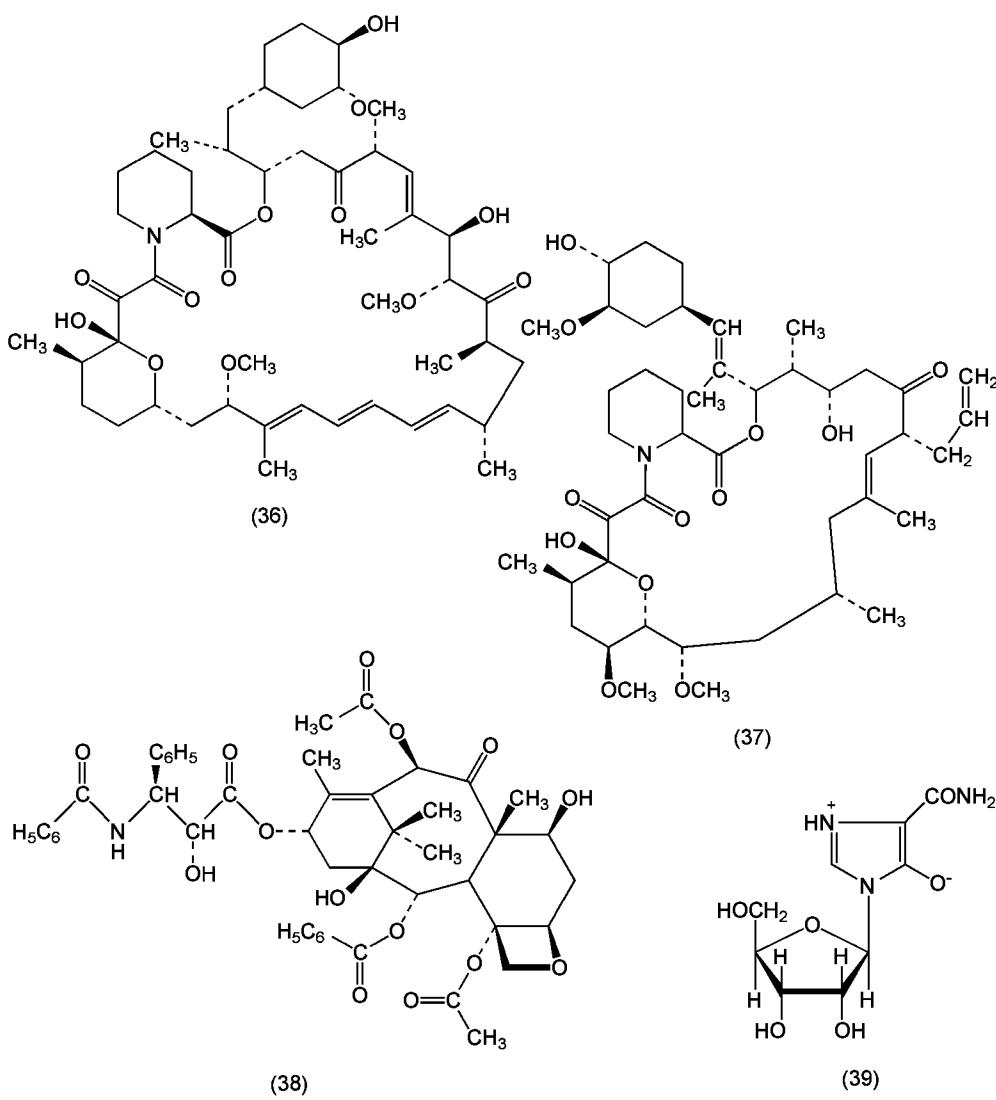


Table 2. Traditional Antirheumatic Drugs Having Immunological Activity^a

Compound	CAS Registry Number	Molecular formula	Structure number
prednisone	[53-03-2]	C ₂₁ H ₂ O ₅	(18)
tenidap ^b	[120210-48-2]	C ₁₄ H ₉ ClN ₂ O ₃ S	(12)
SKF 86002	[72873-74-6]	C ₁ H ₁₂ FN ₃ S	(20)
E-5110		C ₂₀ H ₂₉ NO ₃	(21)
auranofin	[34031-32-8]	C ₂₀ H ₃₄ AuO ₉ PS	(22)
aurothioglucose	[12192-57-3]	C H ₁₁ AuO ₅ S	(23)
gold sodium thiomalate	[24145-43-5]	C ₄ H ₅ AuO ₄ S	(24)
chloroquine	[54-05-7]	C ₁₈ H ₂ ClN ₃	(25, <i>R – H</i>)
mefloquine	[53230-10-7]	C ₁₇ H ₁ F N ₂ O	(26)
tetrandrine	[518-34-3]	C ₃₈ H ₄₂ N ₂ O	(27)
general IL-1 inhibitor			(28)
D-penicillamine	[52-67-5]	C ₅ H ₁₁ NO ₂ S	(29)
sulfasalazine	[599-79-1]	C ₁₈ H ₁₄ N ₄ O ₅ S	(30)

^a See Figure 2.
^b This drug is considered an immunomodulator. The structure appears in Figure 1.

Table 3. Immunosuppressive Drugs^a

Compound	CAS Registry Number	Molecular formula	Structure number
<i>Synthetic products</i>			
methotrexate	[59-05-2]	C ₂₀ H ₂₂ N ₈ O ₅	(19)
azathioprine	[446-86-6]	C ₉ H ₇ N ₇ O ₂ S	(31)
cyclophosphamide	[50-18-0]	C ₇ H ₁₅ Cl ₂ N ₂ O ₂ P	(32)
mycophenolate	[115007-34-6]	C ₂₃ H ₃₁ NO ₇	(33)
brequinar	[96187-53-0]	C ₂₃ H ₁₅ F ₂ NO ₂	(34)

deoxyspergualin			
		<i>Natural products</i>	
cyclosporin A	[59865-13-3]	$C_{22}H_{111}N_{11}O_{12}$	(35)
sirolimus	[53123-88-9]	$C_{51}H_{79}NO_{13}$	(36)
FK 506	[104987-11-3]	$C_{44}H_{99}NO_{12}$	(37)
taxol	[33069-62-4]	$C_{47}H_{51}NO_{14}$	(38)
mizoribine	[50924-49-7]	$C_9H_{13}N_5O$	(39)

^a See Figures 3 and 4.

Corticosteroids (60) are used to treat some patients with severe progressive RA. Low dose prednisone (**18**) may be a better alternative to second-line therapy for the elderly. In younger patients, disease control may be desired temporarily until second-line drugs, with a slower onset, can provide sufficient control. For patients who cannot tolerate NSAIDs or have severe systemic manifestations of RA, such a pericarditis or vasculitis, prednisone may be helpful. Intraarticular injections of corticosteroids are often helpful in treating acute inflammation of RA joints. There are many adverse side effects of corticosteroids. These include osteoporosis, cataracts, poor healing, gastrointestinal bleeding, hyperglycemia, hypertension, and increased infection. Gastrointestinal bleeding and osteoporosis may be more severe in the elderly.

A new generation of antiinflammatory agents having immunosuppressive activity has been developed. The appearance of preclinical and clinical reports suggest that these are near entry to the pharmaceutical market. For example, tenidap (CP-66,248) (**12**) has been demonstrated to inhibit IL-1 production from human peripheral blood monocytes in culture (55). Clinically, IL-1 in synovial fluids of arthritic patients was reduced following treatment with tenidap. Patients with rheumatoid or osteoarthritis, when treated with tenidap, showed clinical improvement (57,58). In addition to its immunological effects, tenidap also has an antiinflammatory profile similar to the classical NSAIDs (59). Other synthetic inhibitors of IL-1 production are SKF 86002 (**20**) and E-5110 (**21**) (55).

The hydroxy derivative of chloroquine (**25**, $R = H$), hydroxychloroquine [118-42-3] (**25**, $R = OH$), an antimalarial, has been shown to be effective for RA (60). Serious side effects are rare, but can include retinal damage and various skin, central nervous system, and bone marrow toxicities. Vision problems should be monitored at six-month intervals, and the drug discontinued at the first signs of renal toxicity. Higher doses cause greater risk. Mefloquine (**26**), another antimalarial compound of the quinoline series, and hydroxychloroquine have been shown to prevent the proteoglycan degradation in cartilage induced by IL-1 (55). Tetrandrine (**27**), a bisbenzylisoquinoline, can influence monocyte–macrophage functions, and has been shown to inhibit IL-1 production from human monocytes. A general generic structure (**28**) for IL-1 inhibitory activity has been postulated (61).

Gold compounds can be effective for active RA and may delay or prevent erosive progression of joints in some patients. Gold sodium thiomalate (**24**), and aurothioglucose (**23**) are available as injectable preparations. An oral preparation is also available, but is less effective and frequently causes diarrhea. Injectable gold is administered in 10-mg amounts as a test dose, followed by 25 mg once weekly for two weeks, then 50 mg weekly for 20 weeks. If a response occurs, treatment intervals may be lengthened to every two weeks, then three weeks, and then monthly. Patients who respond usually remain at least on monthly therapy. Discontinuation of gold therapy maintenance may result in recurrence of arthritic symptoms, which may not remit even with reinstitution of gold therapy. Auranofin (**22**), administered in 3-mg amounts twice daily or 6 mg once daily, should be continued for at least six months, assuming a favorable response.

Adverse side effects of gold treatments include stomatitis, rash, and proteinuria. Complete blood counts and urinalysis should be performed before each or every other injection of gold compounds. Pruritic skin rash and stomatitis are more common adverse effects that may resolve, if therapy is withheld for a few weeks and then restarted cautiously at a lower dose. Oral gold causes less mucocutaneous, bone marrow, and renal toxicity than injectable gold, but more diarrhea and other gastrointestinal reactions appear.

Penicillamine (**29**) can be effective in patients with refractory RA and may delay progression of erosions, but adverse effects limit its usefulness. The most common adverse side effects for penicillamine are similar to those of parenteral gold therapy, ie, pruritic rash, protein uria, leukopenia, and thrombocytopenia. Decreased or altered taste sensation is a relatively common adverse effect for penicillamine. A monthly blood count, platelet count, and urinalysis are recommended, and also hepatic and renal function should be periodically monitored. Penicillamine is teratogenic and should not be used during pregnancy.

Sulfasalazine (**30**), used for many years for the treatment of inflammatory bowel disease (60), is increasingly used for RA. Sulfasalazine appears to be as effective as injectable gold and is better tolerated. Other studies (60) showed sulfasalazine was as effective as penicillamine, but not as toxic. Sulfasalazine may be more effective than hydroxychloroquine in preventing progression of joint erosion. Adverse side effects include gastrointestinal disturbances and rash, but hepatitis and blood dyscrasias are rare. Enteric coating decreases the gastrointestinal toxicity. Monitoring for hepatitis and bone marrow suppression are recommended every 2–3 weeks during the first three months of treatment and less frequently thereafter.

Interferons (IFNs) (52,53), a family of species-specific vertebrate proteins, confer nonspecific resistance to a broad range of viral infections, affect cell proliferation, and modulate immune responses. All three principal interferons, α -interferon (IFN- α) produced by blood leucocytes, β -interferon (IFN- β) by fibroblasts, and γ -interferon (IFN- γ) by lymphocytes, also have antiviral activity. The ability of interferons to inhibit growth of transplantable and carcinogen-induced tumor led to research showing the direct antiproliferative and indirect immune-mediated antitumor activities (see CHEMOTHERAPEUTICS, ANTICANCER). IFNs have been found to be efficacious in certain malignancies and viral infections, eg, hairy cell leukemia (85° response) and basal cell carcinoma (86° response). However, the interferons do have adverse side effects (54).

The gene encoding IFN- γ has been cloned and large quantities of recombinant γ -interferon (r-IFN- γ) are available for clinical testing. Preclinical *in vivo* testing of r-IFN- γ detected no significant differences between the recombinant and natural IFNs. The lack of glycosylation does not affect the plasma half-life of r-IFN- γ . Additionally, r-IFN- γ has important and significant *in vivo* biological effects.

Clinical trials for r-IFN- γ in RA indicated that the drug is well tolerated (52). Consistent improvement in tender and swollen joint scores was observed, but a large number of patients were needed in the trial to show statistical significance for r-IFN- γ treatment. In certain individuals, responses were remarkable. An additive effect between r-IFN- γ and penicillamine was detected. Efficacy was lower when r-IFN- γ was combined with gold therapy. Research is continuing.

Immunosuppressive agents (see Table 3), such as methotrexate (**19**) and azathioprine (**31**), are used increasingly in management of RA. These can suppress inflammation and thus allow a decrease in the doses of corticosteroids used. Oral methotrexate (**19**), a potent immunosuppressive, can be very effective for stopping the erosive changes associated with rheumatoid arthritis (50). The antirheumatic effect of (**19**) weekly, in low doses, is often apparent within 4–6 weeks of treatment and some rheumatologists consider this compound a first-line drug. Although methotrexate is often well tolerated, the drug can cause anorexia, nausea, vomiting, abdominal cramps, hepatic toxicity, bone marrow suppression, and rarely, hepatic fibrosis. Allergic pneumonitis, often severe, characterized by dry cough, fever, breathlessness, and hypoxia, occurs in 1–4° of RA patients taking methotrexate. Infections, eg, herpes zoster and *Pneumocystis carinii*, may also be more common in patients taking methotrexate.

Percutaneous liver biopsy after each 1.5 g of total accumulated methotrexate dosage to detect hepatic fibrosis or cirrhosis not reliably predicted by serum aminotransferase tests are recommended (1,50). Concurrent use of NSAIDs may increase toxicity of methotrexate, although toxicity may be avoided if the drugs are separated by 12 h.

Azathioprine (**31**), a purine analogue immunosuppressive drug (51), can be as effective as gold and penicillamine in patients with refractory RA. Adverse side effects include nausea, vomiting, abdominal pain, and hepatitis. Reversible bone marrow depression can occur, but severe toxicity is uncommon with the dosage used for RA. A complete blood count, serum aminotransferase, bilirubin, and alkaline phosphatase determinations are recommended every 2–4 weeks, until the disease is under control, or adverse side effects such as cytopenia appear. Azathioprine-treated patients may have an added risk of malignancy. The drug should not be used during pregnancy.

Initially, the immunosuppressive agents, such as cyclophosphamide (32), azathioprine, and methotrexate, were developed to inhibit malignant cell proliferation. The immunosuppressant activity was discovered later and these agents were then applied to treat autoimmune diseases, where patients did not respond to high doses of steroids (51). The potential side effects associated with these agents have encouraged the search for unique immunosuppressants having more acceptable safety and efficacy profiles (62). Future approaches need to incorporate early treatment with immunotherapy (63).

Cancer. Cancer is a cellular malignancy characterized by loss of normal controls resulting in unregulated growth, lack of differentiation, and the ability to invade local tissues and metastasize. Most cancers are potentially curable, if detected at an early enough stage. The ideal antineoplastic agent would destroy cancer cells without adverse effects or toxicities to normal cells. No such drug exists.

Modern cancer therapy has been primarily dependent upon surgery, radiotherapy, chemotherapy, and hormonal therapy (72) (see CHEMOTHERAPEUTICS, ANTICANCER; HORMONES; RADIOPHARMACEUTICALS). Chemotherapeutic agents may be able to retard the rate of growth, but are unable to eradicate the entire population of neoplastic cells without significant destruction of normal host tissue. This serious side effect limits general use. More recently, the immunotherapeutic approach to cancer has involved modification and exploitation of the cellular and molecular mechanisms in host defense, regulation of tissue proliferation, tissue differentiation, and tissue survival. The results have been more than encouraging.

Natural products that are immune stimulants, eg, *Bacillus calmette guerin* (BCG), the purified protein derivative (PPD) of tuberculin, and crude lymphokine preparations, have been used to attempt regression of numerous primary or metastatic skin malignancies. This modality has also been used to treat basal cell carcinoma, mycosis fungoids, lymphangiosarcoma, reticulum cell sarcoma, and breast cancer (76). However, this local response appears to be selective for tumor cells and is relatively sparing of the normal adjacent tissues. Active nonspecific immune stimulation can be accomplished, not only using BCG, but also other intact organisms, eg, *Corynebacterium parvum*, *Bordetella pertussis*, penicillin-inactivated *Streptococcus* OK 432 endotoxin, yeast, and products extracted from yeast. It has been speculated that these nonspecific approaches may enhance specific immune responses to tumor antigens or tumor cells. However, the success of such approaches has not been reproducible.

Modern clinical immunotherapy of cancer (65) began in the 1960s with treatment of localized skin malignancies. By the 1980s human interferons were known to modulate the expression and shedding of HLA and other melanoma-associated antigens. The availability of unlimited quantities of purified interferon through recombinant DNA technology led to in-depth studies and the extensive use of interferons in the clinic. In the early 1990s immunologic treatment modalities for primary or metastatic malignant melanoma include interferons as single agents, or in combination with other cytotoxic drugs, monoclonal antibodies, and melanoma vaccines.

The concept that a weak or suppressed host immune response may be overridden by active immunization appears to be valid. Vaccines of autologous tumor extracts have been used to treat patients with malignant melanoma. Interesting and significant progress has been made in the treatment of neoplasia using interferons and interdeukin-2. A particularly effective therapy for tumor patients is the infusion of lymphocyte-activated killer (LAK) cells, produced *in vitro* by the incubation of peripheral blood mononuclear cells (PBMc) from cancer patients with interdeukin-2 (IL-2) (69,70). These IL-2-induced LAK cells have been shown to lyse autologous and allogeneic tumor cells *in vitro* and *in vivo* (71,72).

Cytokines, eg, interferons, interdeukins, tumor necrosis factor (TNF), and certain growth factors, could have antitumor activity directly, or may modulate cellular mechanisms of antitumor activity (2). Cytokines may be used to influence the proliferation and differentiation of T-cells, B-cells, macrophage-monocyte, myeloid, or other hematopoietic cells. Alternatively, the induction of interferon release may represent an important approach for synthetic-medicinal chemistry, to search for effective antiinflammatory and antifibrotic agents. Inducers of interferon release may also be useful for lepromatous leprosy and chronic granulomatous disease. The potential cytokine and cytokine-related therapeutic approaches to treatment of disease are summarized in Table 4. A combination of cytokines is a feasible modality for treatment of immunologically related diseases; however, there are dangers inherent in such an approach, as shown by the induction of lethal disseminated intravascular coagulation in mice administered TNF- α and IFN- γ .

Table 4. Cytokine and Cytokine-Related Therapeutic Approaches to Disease^a

Cytokine antibody ^b	Disease target	Species tested
IL-1	radiation cytotoxic injury bacterial infection	rodent
TNF- α	autoimmune lupus nephritis	rodent
TNF- β	tumor destruction	rodent
INFs	antiinflammatory immunoregulation	rodent and human
INF- α , β , γ	tumor destruction, tumor and lymphocyte-induced angiogenesis	human rodent
INF- γ	rheumatoid arthritis lepromatous leprosy chronic granulomatous disease	human human human
IFN inducers		
poly I:C	fibrosis; transplantation	rodent
tilorone	adjuvant arthritis DTH granuloma	rodent rodent
IL-2 + LAK cell or tumor infiltrating lymphocyte	tumor destruction	rodent and human
GM-CSF, G-CSF, M-CSF, multi-CSF	cytotoxic injury; bone marrow transplantation; myelodysplastic syndromes; AIDS neutropenia	rodent and human
CSF-1 (M-CSF)	tumor destruction	rodent
basic FGF (bovine)	cartilage repair	rabbit
GM-CSF Ab + IL-3 Ab	cerebral malaria	rodent
IL-4 Ab	allergy; parasitic infection	rodent

^a Ref. 77.
^b Ab = Antibody; IL = interleukin; TNF = tumornecrosis factor; INF = interferon; LAK = lymphocyte-activated killer; CSF = colony stimulating factors; and FGF = fibroblast growth factor.

The active immunotherapeutic approach is specific and based on the premise that tumor antigens are immunogenic and the host is sufficiently immunocompetent to mount an effective immune response to an autologous tumor. Theoretically, a weak or suppressed host immune system that had allowed the formation of a tumor may be overridden by active immunization or immunostimulation. In practice, vaccines composed of so-called autologous tumor extracts have been used to treat patients with malignant melanoma (73), and purified melanoma tumor-associated antigens have been used to elicit antibody responses in melanoma patients (74).

The conjugation of monoclonal antibodies (MoAbs) to radioisotopes, chemotherapeutic agents, and protein toxins has also been given consideration (65). Large amounts of human MoAbs can be produced by biotechnological means.

The combination of different therapeutic modalities, surgery, chemotherapy, or radiation, as well as a combination of different chemotherapeutic

and biotechnology-engineered agents, have been used in the treatment of a number of neoplasms and led to higher rates of response (75). Investigations using biological response modifiers have produced very encouraging clinical benefits (48,64), as seen in therapy of breast cancer, myeloma, non-Hodgkin's lymphomas, hairy cell leukemia, essential thrombocythemia, and renal cell carcinoma. Taxol (38) (see Table 3), a natural product initially isolated from the bark of the Pacific Yew tree, *Taxus brevifolia*, is an antileukemic and antitumor agent (67). It promotes assembly of microtubules and inhibits the disassembly process. The total synthesis of taxol has been achieved. Partial synthesis of derivatives using products from renewable parts of taxol sources allows preparation of a taxol-related series of antitumor agents.

Another natural product, mizoribine (39), a nucleoside antibiotic produced by the fungus *Eupenicillium brefeldianum*, has cytotoxic and immunosuppressive activity. It has been evaluated for use in renal transplantation and neoplasia (68).

Transplantation. Advances in surgical techniques and the availability of selective immunosuppressants, along with careful patient selection and proper post-surgery management, are factors in making transplantation the treatment of choice for organ failure. During the 1980s, the discovery of cyclosporin A (35) played a role in the significant increase in graft survival (78,79) (see ANTIBIOTICS, PEPTIDES). But the use of transplants is still limited because of the acute immune rejection phenomenon (host vs graft reaction, or HVGR) which may destroy the transplanted tissue within days to months after the surgery. Safe and effective immunotherapeutic agents, used to control and regulate the HVGR, are crucial to the development of better transplantation methods. The primary goal is to achieve selective suppression of the recipient's immune response to the foreign antigens in the graft, ie, to attain specific immunological tolerance only to specific antigens.

Although there is no reliable method as of this writing for induction of Ag-specific unresponsiveness, some degree of tolerance has been observed by use of nonspecific immunosuppressive therapy. This conclusion is supported by a decrease in the frequency of precursor T-cells reactive with graft HLA Ags in long-term recipients of organ transplants.

Nonspecific immunosuppressive therapy in an adult patient is usually through cyclosporin (35), started intravenously at the time of transplantation, and given orally once feeding is tolerated. Typically, methylprednisone is started also at the time of transplantation, then reduced to a maintenance dose. Azathioprine (31) may also be used in conjunction with the prednisone to achieve adequate immunosuppression. Whereas the objective of immunosuppression is to protect the transplant, general or excessive immunosuppression may lead to undesirable complications, eg, opportunistic infections and potential malignancies. These adverse effects could be avoided if selective immunosuppression could be achieved. Suspected rejection episodes are treated with intravenous corticosteroids. Steroid-resistant rejection may be treated with monoclonal antibodies (78,79) such as Muromonab-CD3, specific for the T3-receptor on human T-cells. Alternatively, antithymocyte globulin (ATG) may be used against both B- and T-cells.

A number of fungal immunosuppressives have been isolated from fermentation broths and demonstrated to have immunotherapeutic efficacy. Other than cyclosporin (35), two fungal metabolites, sirolimus (36), previously known as rapamycin (80), and FK-506 (37) (81) are in various stages of development (see ANTIBIOTICS, MACROLIDES).

Cyclosporin (35), a drug of choice in transplantation, can inhibit synthesis and release of IL-2 and other cytokines that result from interactions between an antigen presenting cell and a T-cell. Significant therapeutic effects have also been reported in patients with RA and associated autoimmune disease. Potential adverse side effects may be decreased by careful monitoring of cyclosporin blood levels, renal function parameters, liver enzymes, and blood pressure. Creative approaches to dosing, eg, every other day administrations or individualized dosing, may be very beneficial to prevent toxic effects. Upon demonstration of efficacy, gradual reduction of the cyclosporin dose reduces the frequency of cyclosporin-manifested side effects (82).

Immunosuppression induced by sirolimus (36) appears to be mediated by a mechanism distinctly different from that of either cyclosporin or FK-506. Sirolimus markedly suppresses IL-2 or IL-4-driven T-cell proliferation. The preclinical studies suggest that sirolimus is a potent immunosuppressive agent in transplantation and autoimmune disease models. The clinical potential of this agent depends on its toxicity profile (80).

FK-506 (37) interferes with IL-2 synthesis and release and has a cyclosporin-like profile, but is considerably more potent *in vitro*. IC₅₀ values are approximately 100-fold lower. This neutral macrolide suppresses the mixed lymphocyte reaction; T-cell proliferation; generation of cytotoxic T-cells; production of T-cell derived soluble mediators, such as IL-2, IL-3, and γ-IFN; and IL-2 receptor expression (83). Structurally, FK-506 is similar to sirolimus. Mycophenolate mofetil (33), brequinar (34), and deoxyspergualin are in various phases of clinical evaluation. Identification of therapeutic efficacy and safety are important factors in the determination of their utility as immunosuppressive agents.

BIBLIOGRAPHY

"Immunotherapeutic Agents" in *ECT* 3rd ed., Vol. 13, pp. 167–182, by S. Wong, Boehringer-Ingelheim Ltd.

1. R. Berkow and A. J. Flether, eds., *The Merck Manual of Diagnosis and Therapy*, Merck & Co., Inc. Rahway, N.J., 1992, pp. 4–219.
2. C. J. Dunn in E. S. Kimball, ed., *Cytokines and Inflammation*, CRC Press, Inc., Boca Raton, Fla., 1991, pp. 1–33.
3. A. Shaw in Ref. 2, pp. 35–59.
4. M. M. Bertagnolli in *Cytokines and T Lymphocytes*, R. G. Landes Co. Publishers, Austin, Tex., 1993, p. 115.
5. A. W. Lohse and I. R. Cohen, in N. R. Farid, ed., *The Immunogenetics of Autoimmune Diseases*, Vol. I, CRC Press, Boca Raton, Fla., 1991, pp. 159–165; A. Ben-Nun and I. R. Cohen. *J. Immunol.* **129**, 303–308 (1982).
6. Ref. 1, pp. 303–313.
7. Ref. 1, pp. 304–310.
8. J. A. Tami, *Pharmacy Immunology*, Special Issue, May 1989.
9. Ref. 1, pp. 303–304.
10. A. J. Bint, J. Oxford, and P. J. Daly, eds., *AIDS and AIDS-related Infections: Current Strategies for Prevention and Therapy*, Academic Press Ltd., San Diego, Calif., 1989, p. 137.
11. J. Mann, D. J. M. Tarantola, and T. W. Netter, eds., *AIDS in the World—A Global Report*, Harvard University Press, Cambridge, Mass., 1992, p. 1037.
12. J. Weber, in Ref. 10, pp. 1–9.
13. M. A. Sande, *J. Antimicrob. Chemother.* **23**, 63–65 (1989).
14. Ref. 1, 1992, pp. 311–318.
15. J. S. Oxford, A. R. M. Coates, D. Y. Sia, K. Brown, and S. Asad, in Ref. 10, pp. 9–28.
16. D. J. Jeffries, in Ref. 10, pp. 29–34.
17. E. De Clercq, in Ref. 10, pp. 35–46.
18. D. Lednicer and K. M. Snader, in H. Wagner and N. R. Farnsworth, eds., *Economic and Medicinal Plant Research*, Vol. 5, Academic Press, Orlando, Fla., 1991, pp. 1–20.
19. C.-T. Che, in Ref. 18, pp. 167–251.
20. T. Okuda, T. Yoshida, and T. Hatano, in Ref. 18, pp. 130–166.
21. J. P. Devlin, C. Carter, and C. A. Homon, in Ref. 18, pp. 84–109.
22. R. Bauer and H. Wagner, in Ref. 18, pp. 253–321.
23. S. Budavari, M. J. O'neil, A. Smith, and P. E. Heckelman, eds., *The Merck Index, An Encyclopedia of Chemicals, Drugs, and Biologicals*, 11th ed., Merck & Co., Inc., Rahway, N.J., 1989, p. 1912.
24. P. T. Daniel, F. Falcioni, A. U. J. Berg, and P. A. Berg, K. Masek ed., in *Immunomodulators—Current Problems in Testing and Evaluation of Experimental and Clinical Effects of Immunomodulators*, J. R. Prous Science, Barcelona, 1986, pp. 139–145.
25. N. Brattig, G. J. Diao, and P. A. Berg, *Int. J. Immunopharmacol.* **6**, 205–215 (1984).

26. G. Rauch, in Ref. 24, pp. 147–150.

27. F. Demeulenaere and co-workers, *Proc. Intl. Workshop (Royal Soc. Med.)*, Crans-Montana, Switzerland, 135–141 (1981).

28. Z. Rudkowski, E. Saraczynska, and B. Muller, *Aktuelle Ergebnisse aus Experiment und Klinik*, Workshop Bad Kreuznach, 1983, pp. 89–110.

29. K. Kino and co-workers, *J. Biol. Chem.* **264**, 472–478 (1989).

30. S. Tanaka and co-workers, *JBC* **264**, 16372–16377 (1989).

31. M. Ferencik and co-workers in Ref. 4, pp. 163–166.

32. M. L. Patchen, T. L. MacVittie, and I. Brook, in Ref. 24, pp. 151–155.

33. N. R. DiIuzio, *Trends Pharm. Sci.* **4**, 344–347 (1983).

34. W. Wooles and co-workers, *J. Reticuloendothel. Soc.* **1**, 160–169 (1964).

35. W. Wooles and co-workers, *Science* **142**, 1078–1080 (1963).

36. W. Wooles and co-workers, *Proc. Soc. Exp. Biol. Med.* **115**, 756–759 (1964).

37. C. Burgaleta and co-workers, *Cancer Res.* **37**, 1739–1742 (1977).

38. M. Patchen and co-workers, *Exp. Hemat.* **8**, 409–422 (1981).

39. M. Patchen and co-workers, *Immunopharmacol.* **5**, 303–313 (1983).

40. N. DiIuzio and co-workers, *Cancer Treat. Rep.* **62**, 1857–1866 (1978).

41. D. Williams and co-workers, *Hepatology* **5**, 198–206 (1985).

42. Ref. 23, p. Ther-25.

43. A. J. Glasky and J. Gordon, in Ref. 24, pp. 35–40.

44. G. Renoux, in Ref. 24, pp. 45–50.

45. E. Lemarie and co-workers in Ref. 24, pp. 51–54.

46. J. M. Lang and co-workers, *Lancet* **2**, 1066 (1985).

47. J. M. Lang and co-workers, *Int. J. Immunopharmacol.* **7**, 337 (1985).

48. R. Renoux and co-workers, in Th. Klein and co-workers, *Biological Response Modifiers in Human Oncology and Immunology*, Plenum Publishing Corp., New York, 1983, pp. 223–239.

49. J. Corke and R. Renoux, *First World Conference on Inflammation*, Venice, Italy Apr. 1984.

50. J. M. Kremer, in D. E. Frust and M. E. Weiblat, eds., *Immunomodulators in the Rheumatic Diseases*, Marcel Dekker, Inc., New York, 1990, pp. 1–21.

51. G. W. Cannon and H. J. Williams, in Ref. 50, pp. 23–59.

52. J. D. Schindler and D. V. Faller, in Ref. 50, pp. 185–222.

53. Z. Estrov, R. Kurzrock, and M. Talpaz in Z. Estrov, ed., *Interferons*, 1993, p. 120.

54. G. G. Steinman, F. Rosenkaimer, and G. Leitz in G. W. Richter, K. Solez, and B. Riffel, eds., *Internat. Rev. Exptl. Path.* **34** (Pt. B), 193–207 (1993).

55. F. J. Persico, in Ref. 2, pp. 59–87.

56. A. S. Fauci, *J. Immunopharmacol.* **1**, 1–25 (1978).

57. P. Katz, A. P. Borger, and L. D. Loose, *Arth. Rheum.* **31**, S52, 1988.

58. J. S. Davis, L. Loose, and A. P. Borger, *Arth. Rheum.* **31**, S72, 1988.

59. I. G. Otterness, T. J. Carty, and L. Loose, in A. J. Lewis and N. R. Ackerman, eds., *Therapeutic Approaches to Inflammatory Diseases*, Elsevier Science Publishing Co., Inc., New York, 1991, pp.

60. *Rheumatoid Arthritis (RA)* in Ref. 1, pp. 1305–1312.

61. **U. S. Pat. 4778806** (1988), P. E. Bender, D. E. Griswold, N. Hanna, and J. C. Lee.

62. J. G. Lombardino and I. G. Otterness, *J. Med. Chem.* **24**, 830–834 (1981).

63. R. I. Fox and co-workers, in Ref. 50, pp. 223–256.

64. M. A. Chirigos in Ref. 48, pp. 27–33.

65. C. W. Taylor and E. M. Hersh, *Immunopharmacol. Rev.* **1**, 89–157 (1990).

66. *Antineoplastic Drugs*, in Ref. 1, Chapt. 103, p. 1277.

67. *Taxol* in Ref. 23, p. 1435.

68. *Mizoribin* in Ref. 23, p. 980.

69. E. Grimm and co-workers, *J. Exp. Med.* **158**, 1356–1361 (1983).

70. E. Grim and co-workers, *J. Exp. Med.* **157**, 884–897 (1983).

71. J. J. Mule and co-workers, *Science* **225**, 1487–1489 (1984).

72. S. A. Rosenberg and co-workers, *N. Engl. J. Med.* **313**, 1485–1492 (1985).

73. D. Berd, H. C. Maguire, Jr., and M. J. Mastrangelo, *Cancer Res.* **46**, 2572–2577 (1986).

74. P. O. Livingston and co-workers, *Proc. Nat. Acad. Sci. USA* **84**, 2911–2915 (1987).

75. Ref. 1, pp. 1285–1287.

76. E. Klein and co-workers, *Med. Clin. North Amer.* **60**, 389–418 (1976).

77. C. J. Dunn in Ref. 2, pp. 1–33.

78. Ref. 1, pp. 346–363.

79. A. Kishore and J. G. Barbaccia, *Pharm. Immunol.* **3**, 1 (1989).

80. S. N. Sehgal and J. Y. Chang, *SIM Indust. News* **41**, 230–234 (1991).

81. T. Kino and co-workers, *J. Biotics* **40**, 1249–1265 (1987).

82. A. Kowal, J. H. Carstens, Jr., and J. J. Schnitzer, in Ref. 52, pp. 61–98.

83. I. G. Otterness and M. L. Bliven, in S. C. Gilman and T. J. Rogers, eds., *Immunopharmacology*, Caldwell, N.J., 1989, pp. 313–342.

Stewart Wong
Jing Xing Health and Safety Resources, Inc.

INCENDIARIES.

See CHEMICALS IN WAR; PYROTECHNICS.

INCINERATORS

Municipalities and industries are encouraged to reduce waste generation. Nevertheless, even under maximum use of source reduction and recycling (qv), significant quantities of waste continue to be generated (see WASTE, INDUSTRIAL). As of this writing, high temperature incineration is the preferred technology for managing these wastes (1–3). Properly designed incinerators have the capability to destroy nearly 100° of all types of liquid organic wastes and an estimated 60° of solid wastes. However, as shown in Table 1, incineration is extremely limited in the United States. Roughly 60° of total U.S. wastes generated annually are classified as hazardous waste and less than 0.5° is incinerated (see WASTE TREATMENT, HAZARDOUS WASTE). About 15° of the municipal solid waste generated is disposed of in incineration systems and effectively all of the medical waste.

Table 1. Summary of U.S. Waste Generated and Incinerated

Type	Waste, 10 ⁶ t yr		Number of operating incinerators
	Generated	Incinerated	
hazardous ^a	249.3	1.3	171
municipal ^b	163	26	168
medical ^c	0.4	0.4	6850 ^d

^a 1987 Estimate (4).

^b 1988 Estimate (5).

^c 1990 Estimate (6).

^d 1991 Estimate (7).

High (\$300–1900 t) capital investment and operating costs help discourage use of incineration for hazardous wastes. These costs for incinerators are well above those for alternative treatment methods (8) such as biological treatment, \$60–770 t; landfill in drums, \$260–740 t or in bulk, \$90–150 t; or deep well injection, \$20–140 t. A primary contributor to the high operating cost of incinerators is the need for auxiliary fuel, particularly for the disposal of liquid wastes having high water content or solid wastes having low heating value. In addition, gas scrubbers can consume large quantities of chemicals, especially if chemical addition is not carefully controlled. Moreover, because incineration systems are typically complex, highly skilled operators are required to ensure efficient and reliable operation.

U.S. Regulations Impacting Design and Operation of Incinerators

U.S. regulations governing the design and operation of incinerators include the Resource Conservation and Recovery Act (RCRA), the Toxic Substances Control Act (TSCA), and the Clean Air Act Amendments of 1990 (CAA). Many states are authorized to regulate hazardous waste and incinerator programs, and state regulations are generally more stringent than federal.

Resource Conservation and Recovery Act. The RCRA, Subtitle C, regulates hazardous waste disposal. It identifies wastes as being hazardous if falling into one or more of the following categories: ignitable, eg, having a flash point less than 60°C; corrosive, eg, having a pH less than 2 or greater than 12.5; reactive, eg, reacting violently when mixed with water; toxic, as determined by the toxicity characteristic leaching procedure (9) (see TOXICOLOGY), or listed in Subtitle C as a hazardous waste from nonspecific sources, industry-specific sources, or as an acute hazardous or toxic waste.

RCRA incinerator regulations include administrative as well as performance standards. Administrative standards include procedures for waste analysis, inspection of equipment, monitoring, and facility security. Steps needed to meet administrative standards are outlined in the permit application; performance standards are demonstrated during a trial burn. Trial burn operating conditions are included in the permit to assure ongoing compliance with the performance standards. Performance standards include destruction and removal efficiency (DRE), particulate emissions limits, products of incomplete combustion emission limits, metal emission limits, and HCl and Cl emission limits (see EXHAUST CONTROL, INDUSTRIAL).

Destruction and Removal Efficiency. In preparation for a trial burn, the owner prepares an analysis of the waste feed stream. Roughly 400 principal organic hazardous constituents (POHCs) are listed in the *Code of Federal Regulations* (10), and if the waste analysis includes POHCs, these are listed along with their concentrations in the incinerator permit application. A government agency, eg, the U.S. EPA, permit writer selects one or more POHC to be used during the trial burn to demonstrate the incinerator's destruction and removal efficiency (DRE). The writer bases the selection on incineration difficulty, ie, low heating value or high thermal stability, and concentration. Constituents having concentrations less than about 100 ppm are likely not to be selected because of difficulty in analytical detection of such small quantities at 99.99° or greater DRE. During the trial burn, for each POHC selected, the incinerator must demonstrate a DRE of 99.99° or greater for RCRA wastes, and 99.9999° for polychlorinated biphenyl (PCB) or dioxin wastes.

DRE

$$\frac{W_{in}-W_{out}}{W_{out}} \times 100$$

where W_m is the mass feed rate into an incineration system of a given POHC and W_{out} is the mass emission rate of the same POHC at the stack to go into the atmosphere. If ash from a municipal incinerator contains metals or other contaminants in amounts that would cause the ash to be classified as hazardous, it must be treated as a hazardous waste under RCRA, Subtitle C regulations. Figure 1 shows a schematic of an incinerator system.

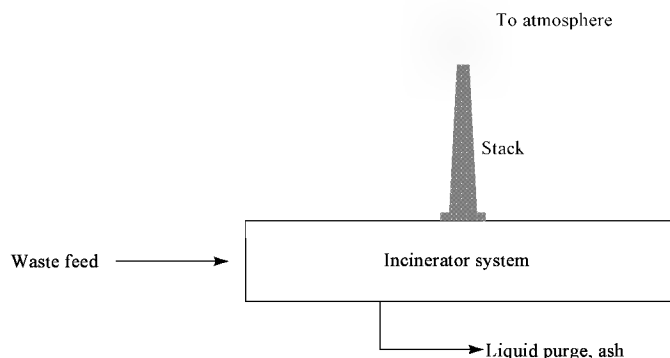


Fig. 1. Schematic of an incinerator system.

Particulate Emission Limits. Particulate emissions, including condensables, must be demonstrated during the burn to be on a dry basis less than 0.18 g m^{-3} (0.08 grain per dry standard cubic foot (gr DSCF)) at the incinerator's stack, or lower depending on state regulations.

Products of Incomplete Combustion Emission Limits. Products of incomplete combustion typically are not directly measured during the trial burn. Instead, levels of carbon monoxide (qv) emissions are used as an indication of combustion efficiency. High combustion efficiencies are assumed to result in acceptable levels of products of incomplete combustion. If carbon monoxide emissions are measured at less than 100 ppmv dry basis, the standard is met. However, if emissions are greater than 100 ppmv, no more than 20 ppmv of total hydrocarbons (qv) are allowed at the incinerator stack during the trial burn.

Metal Emission Limits. Limits for metals, both carcinogenic and noncarcinogenic, are based on an adjusted stack height. Failure to meet these limits requires risk assessments using site specific factors and modeling to establish limits for each metal. The assessments are based on the probability of developing adverse health effects or cancer, based on an inhalation exposure pathway to maximum exposed individuals located near the incinerator (see HAZARD ANALYSIS AND RISK ASSESSMENT).

HCl and Cl Emission Limits. Hydrochloric acid and chlorine must also meet emission limits for an adjusted stack height.

Other Federal Regulations. The TSCA regulates the operation of incinerators disposing of waste containing more than 50 ppm PCBs. Such units must demonstrate a 99.9999 destruction and removal efficiency during a trial burn prior to obtaining an operating permit. The CAA governs emissions of ozone (qv), carbon monoxide, particulate matter, sulfur dioxide, nitrogen oxides, and lead (qv) from incinerators. It is intended to maintain National Ambient Air Quality Standards (NAAQS) for each of these pollutants in attainment areas, and to improve the quality of air in nonattainment areas. This regulation is generally more restrictive than RCRA because it specifies the type of emission control technology to be used when the potential to emit exceeds specified levels (see AIR POLLUTION; AIR POLLUTION CONTROL METHODS).

Solid Waste Incineration

Polymeric or carbonaceous solids are degraded by high temperature. In the presence of oxygen, any carbon, hydrogen, and sulfur are oxidized to CO_2 , H_2O , and SO_2 , respectively. The rate of incineration increases rapidly with temperature. A range of 700–760°C is generally required for combustion, and most general-purpose incinerators operate between 760 and 1100°C. For an incinerator to operate without auxiliary fuel or air preheating, the waste feed or refuse must contain less than 50% moisture or 60% ash, and have more than 25% combustibles.

Atmospheric Conditions. In addition to complete combustion, wastes may be destroyed by treatment at high temperatures either without oxygen (qv) (pyrolysis), using limited oxygen (partial combustion), or in reactive atmospheres (gasification), such as those containing steam (qv), hydrogen (qv), or carbon dioxide (qv).

Pyrolysis. If solids are heated in the absence of oxygen, other than that contained in the feed, organic material breaks down and combustible gas, organic liquids, and char are produced. The yield depends on the properties of the organic solid, temperature, and heating rate. Most of the chemical energy as measured by the heat of combustion originally in the solid appears as chemical energy in the pyrolysis products. To obtain the elevated temperatures required for the reaction, heat must be added to the incinerator combustion zone.

Starved-Air or Partial Combustion. To obtain the temperatures needed for the pyrolysis reaction to occur, a limited amount of oxygen is allowed to enter the combustion zone. This oxygen reacts with the feed or pyrolysis products and releases the needed energy within the combustor. Both pyrolysis and combustion products are obtained. The products leaving the system contain a large amount of chemical energy.

Stoichiometric Air. If the stoichiometric amount of air is added to the combustion zone, the temperature in that zone is controlled by removing heat from the system. No energy leaves the system as chemical energy. The heat is removed by an external means such as generating steam or gas quenching.

Refuse Benefaction. It is extremely difficult to burn and recover useful energy from unsorted municipal waste because of its heterogeneity in size, shape, chemical composition, and heating value. However, preparation of the waste before thermal treatment facilitates burning. Such pretreatment contributes to the front-end costs but reduces furnace costs. The waste is upgraded by separation of the nonorganic fraction and drying, shredding, and densifying solids. These fuels thus prepared are referred to as refuse-derived fuels (RDF) (see FUELS FROM WASTE).

Incinerator Types.

Moving-Grate Incinerators. In this type of incinerator, the waste moves through the furnace on a moving grate. The grate provides support for the refuse, admits the underfire air through openings, transports the refuse from the feed chute to the ash quench, and agitates the bed to bring fresh charge to the surface and expose waste to the oxygen and flame. The refuse fed to the furnace is first dried and preheated by radiation from the hot combustion gases and refractory furnace lining. The refuse, as it is heated further, first pyrolyzes and then ignites. Combustion takes place both in the solid to burn out the residue and in the gas space to burn out the pyrolysis products. Overfire air jets greatly assist the mixing and combustion in the overfire air space.

In U.S. practice, the trend has been from stationary hearths to traveling, rocking, and reciprocating grates (11). Stokers that provide high agitation are preferred even though the agitation increases the particulate loading from the combustion chamber, but not necessarily from the stack, because the air pollution control (APC) system can be designed to handle whatever loading is encountered in the combustion gases. Agitation increases the rate of heat transfer and, hence, combustion in the bed.

The underfire air rate is selected to balance the conflicting demands of high air rates for high burning rates and of low air rates to minimize the particulate loading on the APC system. In many incinerators 60–100% of the stoichiometric air requirements are supplied under the grates. The balance of the air is fed through overfire jets.

There are three types of grate systems encountered most frequently: traveling, reciprocating, and rocking grates. Traveling grates are continuous belt-like conveyors, ie, two or more grates are positioned at different elevations and the solids tumble from one to another to provide agitation. In a reciprocating grate system, movable and fixed sections alternate. The refuse is pushed forward by the fixed sections as the movable sections slide across them. This grate produces more agitation than the traveling grate, but slag formation on the surface can interfere with the reciprocating action. In rocking grates, each section is pivoted. Alternate rows are mechanically rocked to produce an upward and forward motion, thus advancing and agitating the solid waste. This type of grate provides more agitation than the reciprocating grate.

Multichamber Incinerators. The multichamber incinerator is primarily used for commercial and industrial installations. These on-site incinerators, which have capacities up to a few hundred kg h, handle small volumes of materials. The multichamber system provides maximum reduction

of waste with minimum air contamination over a wide range of operating conditions. These are best used for combustible paper (qv), cardboard, wood (qv), foliage, and sweepings.

The combustion process proceeds in two stages: in the primary section the solid phase burns and volatile gases are driven off; in the secondary section, these volatile gases are burned. The combustion of refuse wastes often requires an auxiliary burner to maintain sufficient temperature for complete combustion. Large amounts of excess air, as high as 300%, are frequently used.

Nonconventional Incinerators.

Suspension-Fired Units. The suspension incinerator requires refuse to be shredded to accelerate combustion. After waste shredding, inerts such as metals and glass can be removed to produce refuse derived fuel (RDF). The prepared waste is fed into the furnace by specially designed feeders that distribute it throughout the furnace. Waste is carried in suspension in the combustion air. As a result of the high surface area per unit mass, combustion is rapid. Large particles that do not burn fall onto a moving grate where sufficient time is allowed for combustion.

Slagging Incineration. If maximum volume reduction is desired, the residue remaining after incineration may be liquefied to a slag at temperatures approaching 1650°C. Slagging incineration reduces volume up to 97.5% compared to 50–85% for conventional ashing mode incineration. However, nitrogen oxide formation increases as a result of high temperature, auxiliary fuel may be required, and the operation is likely to be more complex. Furthermore, materials of construction are more expensive.

Starved-Air Incinerators. In order to avoid the entrainment of excessive amounts of particulates from the burning refuse bed, air can be passed more slowly through the grate, producing starved-air incineration. In these units, the combustion of the pyrolysis products leaving the fuel bed is not completed above the burning bed but in a secondary combustion chamber. Sometimes an auxiliary fuel supply is provided in the secondary combustion chamber for periods when wastes having low heating values are burned and during start-up and shutdown. Three or four combustion chambers sometimes follow the primary combustion chamber, with or without flue gas recirculation. The starved-air incinerator has been successfully employed in commercial operation.

Vortex Incinerators. Combustion air is injected tangentially above the burning bed, spirals down through the outside of the bed, and up through the inside. The burning rates reported for this design are only slightly lower, per unit cross section of the incinerator, than the rates encountered in conventional grate units. This incinerator was originally developed for the treatment of paper wastes. For this or any waste yielding a finely subdivided ash, all inert material is carried over with the combustion products to the APC device. Such units, therefore, can be operated continuously without provision for residue removal from the combustion chamber.

Multihearth Furnace. Multihearth furnaces are most often used for incineration of municipal and industrial sludges, and for generation and reactivation of char. The main components of the multihearth are a refractory-lined shell, a central rotating shaft, a series of solid flat hearths, a series of rabble arms having teeth for each hearth, an afterburner (possibly above the top hearth), an exhaust blower, fuel burners, an ash removal system, and a feed system.

The feed is normally introduced to the top hearth where the rabble arms and teeth attached to the central shaft rotate and spiral solids across the hearth to the center, where an opening is provided and the solids drop to the next hearth. The teeth of the rabble arms on the hearth spiral the solids toward the outside to ports that let the solids drop down to the next hearth. Solids continue downward, traversing each hearth until they reach the bottom and the ash is discharged. The primary advantage of this system is the long residence time in the furnace controlled by the speed of the central shaft and pitch of the teeth.

Burners and combustion air ports are located in the walls of the furnace to introduce either heat or air where needed. The air path is countercurrent to the solids, flowing up from the bottom and across each hearth. The top hearth operates at 310–540°C and dries the feed material. The middle hearths, at 760–980°C, provide the combustion of the waste, whereas the bottom hearth cools the ash and preheats the air. If the gas leaving the top hearth is odorous or detrimental to the environment, afterburning is required. The moving parts in such a system are exposed to high temperatures. The hollow central shaft is cooled by passing combustion air through it.

Fluidized-Bed Incinerator. Fluidized-bed incinerators are employed in the paper and petroleum (qv) industries, in the processing of nuclear wastes, and the disposal of sewage sludge. These are quite versatile and can be used for disposal of solids, liquids, and gaseous combustible wastes.

The basic fluid-bed unit consists of a refractory-lined vessel, a perforated plate that supports a bed of granular material and distributes air, a section above the fluid bed referred to as freeboard, an air blower to move air through the unit, a cyclone to remove all but the smallest particulates and return them to the fluid bed, an air preheater for thermal economy, an auxiliary heater for start-up, and a system to move and distribute the feed in the bed. Air is distributed across the cross section of the bed by a distributor to fluidize the granular solids. Over a proper range of airflow velocities, usually 0.8–3.0 m/s, the solids become suspended in the air and move freely through the bed.

The fluidized bed has many desirable characteristics. Because of the movement of the particles, the bed operates isothermally and minimizes hot or cold regions. Large fluctuations in fuel quality are damped out as a result of this thermal capacity. Solid particles are reduced in the bed until they become small and light enough to be carried out of the bed. Bed diameter is limited to about 15 m, and the depth ranges from 0.5 to 3 m. Bed material may be chosen to react with some impurity, eg, SO₂, in the waste to remove it from the gas stream. As a result of the excellent air-to-solid contact, the fluid bed may be operated at low excess air rates. High heat-transfer rates allow large quantities of heat to be removed by a small heat-transfer area in the bed. Fluid beds are not effective in handling materials that contain components with a low ash-melting or softening temperature, since it is often difficult to distribute such materials over the bed cross section.

Rotary Kiln Incinerators. The rotary kiln has been used to incinerate a large variety of liquid and solid industrial wastes. Any liquid capable of being atomized by steam or air can be incinerated, as well as heavy tars, sludges, pallets, and filter cakes. This ability to accept diverse feeds is the outstanding feature of the rotary kiln and, therefore, this type of incinerator is often selected by the chemical and waste treatment industries.

Factors Affecting Destruction of Solid Wastes. The analysis of the evolution and destruction of hydrocarbons during the incineration of solid hazardous wastes involves heat transfer, mass transfer, and reaction kinetics (see HEAT EXCHANGE TECHNOLOGY, HEAT TRANSFER). Figure 2 is a generalized flow chart for the processes experienced by solids during incineration. The key phenomena include the flashing of liquid hydrocarbons; the vaporization, desorption, and stripping of hydrocarbons; the pyrolysis and charring of hydrocarbons; and the oxidation of char. To a certain extent these processes occur in parallel and are common to most thermal treatment processes. Emphasis herein is on applications to the thermal treatment of contaminated soils at temperatures less than 800°C. Slagging kilns are not discussed.

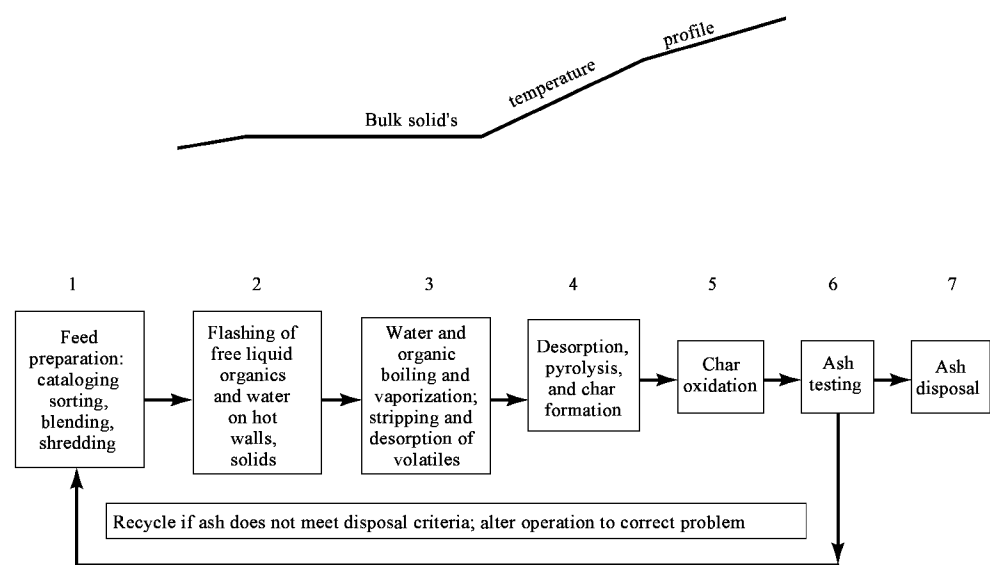


Fig. 2. Generalized process flow chart for the thermal treatment of solid wastes. To a certain extent, steps 2, 3, 4, and 5 always proceed in parallel because of mixing limitations, nonhomogeneities in the waste, and unevenness in its heating.

Vapor Pressures and Adsorption Isotherms. The key variables affecting the rate of destruction of solid wastes are temperature, time, and gas–solid contacting. The effect of temperature on hydrocarbon vaporization rates is readily understood in terms of its effect on liquid and adsorbed hydrocarbon vapor pressures. For liquids, the Clausius-Clapeyron equation yields

$$p = A_1 \exp \left(- \frac{\Delta H_{\text{vap}}}{RT} \right) \tag{1}$$

where A_1 is a constant, p is the partial pressure of the hydrocarbon, T is temperature, R the ideal gas constant, and ΔH_{vap} is the heat of vaporization which is assumed to be independent of temperature. Heats of vaporization for liquid hydrocarbons are typically 30–40 kJ mol (7.2–9.6 kcal mol).

For adsorbed hydrocarbons, the adsorption–desorption process can be thought of as a reaction and the adsorption isotherm as a description of the reaction at equilibrium. For the Langmuir isotherm,

$$\theta = \frac{Kp}{1 + Kp} \tag{2}$$

where K is the adsorption equilibrium constant and θ is the fractional coverage. To find p as a function of T , equation 2 may be rearranged to give

$$Kp = \frac{\theta}{1 - \theta} \quad \text{or} \quad \ln K + \ln p = \text{constant} \tag{3}$$

K and hence p are related to the heat of adsorption (ΔH_{ads}). If ΔH_{ads} is independent of temperature, then at constant θ , where A_2 is a constant,

$$p = A_2 \exp \left(\frac{\Delta H_{\text{ads}}}{RT} \right)_{\theta} \tag{4}$$

Heats of adsorption for hydrocarbons typically range from 20 to 70 kJ mol (4.8 to 16.7 kcal mol). Equations 1 and 4 both indicate that vapor pressures increase exponentially with increasing temperature.

Desorption Rates From Wet and Dry Solids. The presence of water complicates the desorption process in several ways, the most obvious being a large thermal effect associated with its high heat of vaporization. Adsorbed moisture can also dramatically affect hydrocarbon adsorption isotherms by competing for adsorption sites, thus decreasing the number of available sites, and affecting the affinity of the remaining sites for hydrocarbons. In addition, the vaporization of adsorbed and condensed moisture creates a stream of water vapor which strips hydrocarbons from porous solids. These effects are illustrated in Figure 3 which shows adsorbed phase concentrations of *p*-xylene as a function of time at 150°C (12). These results were obtained in a fixed-bed reactor purged with dry N₂. The curves in Figure 3 also indicate the negligible rate of desorption after about 30 minutes. This type of behavior is typical of the desorption of hydrocarbons from soils and other solids (13). To overcome this behavior, higher temperatures are required.

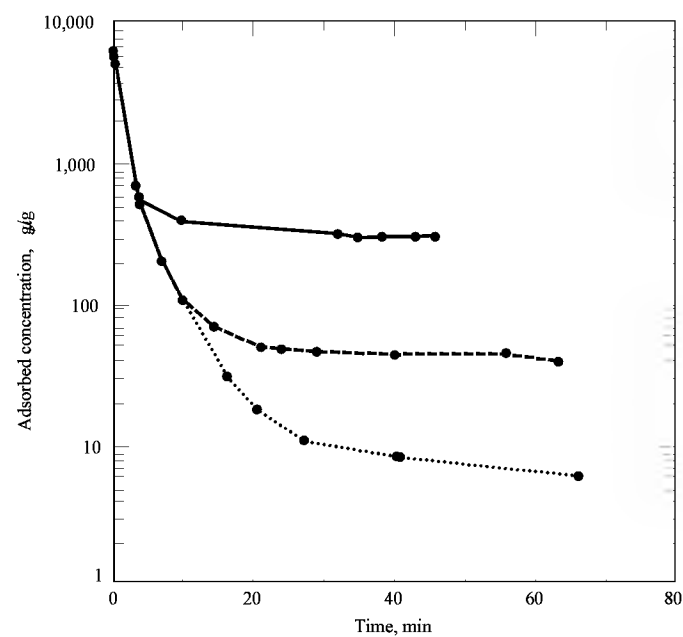


Fig. 3. The effects of moisture on the desorption of *p*-xylene from a clay soil at 150°C where (—●—) represents dry clay; (---●---), clay where water is added after the contaminate; and (···●···), contaminated wet clay from natural moisture (12).

Particle Size and Desorption Rates. Bench-scale reactor studies of the desorption of toluene from single, 2- to 6-mm porous clay particles (14) showed desorption times that increased with the square of the particle radius, suggesting that diffusion controls the rate desorption. Parallel experiments performed in a small, pilot-scale rotary kiln at 300°C showed no effect of clay particle size for diameters ranging from 0.4 to 7 mm. Additional single-particle studies with temperature profiles controlled to match those in the pilot-scale kiln had desorption times that were a factor of 2–3 shorter for the range of sizes studied (15). Hence, at the conditions examined, intraparticle mass transfer controlled the rate of desorption when single particles were involved and interparticle mass transfer controlled in a bed of particles in a rotary kiln. These results apply to full-scale kilns. As particle size is increased, intraparticle resistances to heat and mass transfer eventually begin to dominate.

In general, the desorptive behavior of contaminated soils and solids is so variable that the required thermal treatment conditions are difficult to specify without experimental measurements. Experiments are most easily performed in bench- and pilot-scale facilities. Full-scale behavior can then be predicted using mathematical models of heat transfer, mass transfer, and chemical kinetics.

Bed Mixing in Rotary Kilns. If bed slipping, slagging, and the segregation of solids are neglected, a characteristic time for transverse bed mixing is readily obtained by a combination of geometrical analyses and physical measurements on a slumping bed. The mixing time scale is the approximate time required for the bed to become well mixed. In the slumping regime of bed motion there are two regions in the bed: a surface region and a bulk region. The surface or top plane of the bed periodically slumps. The bulk region moves with the wall of the kiln. Slipping between the bulk region and the kiln wall is detrimental because it reduces the frequency of transverse slumping across the top plane and reduces the rate of mixing. Segregation becomes significant in binary mixtures if the ratio of particle sizes is greater than 6:5 (16).

The geometry of a slumping bed (Fig. 4) is described by the kiln radius, *r*, and the half-angle subtended by the solids (radians), *ψ*. From a material balance on the bed and from the bed's geometry (17), the average time required for a particle to travel through the bulk region and over the top of the bed is given by

$$\tau = \frac{f}{\Omega \sin^2 \psi} \tag{5}$$

where *f* is the fraction of the kiln filled with solid and *Ω* is the kiln rotation rate in units of rev/s. Quantitative measurements of dyed particle weight fractions (18) show that a good estimate of the characteristic time for transverse mixing in slumping beds, *τ_m*, is given by multiplying equation 5 by 3, ie, the ratio of *τ_m* to *τ* is ≈ 3. The degree of incomplete mixing, *M*, is given by

$$M = \exp(-t/\tau_m) \tag{6}$$

for *t* > *τ_m*, *M* = 0.37. The times required to achieve *M* = 0.05 for a kiln rotating at 1 rpm are 60, 85, and 113 s for fill fractions of 0.02, 0.05, and 0.10, respectively. If the rotation rate is reduced to 0.1 rpm the mixing times are 600, 850, and 1130 s, respectively. Lower rotation rates are often chosen because these reduce dust entrainment rates and increase the residence time of the solids. Overloading the kiln can result in incomplete mixing of the bed and limited heating of the solids. Fill fractions much in excess of 0.1 are to be avoided.

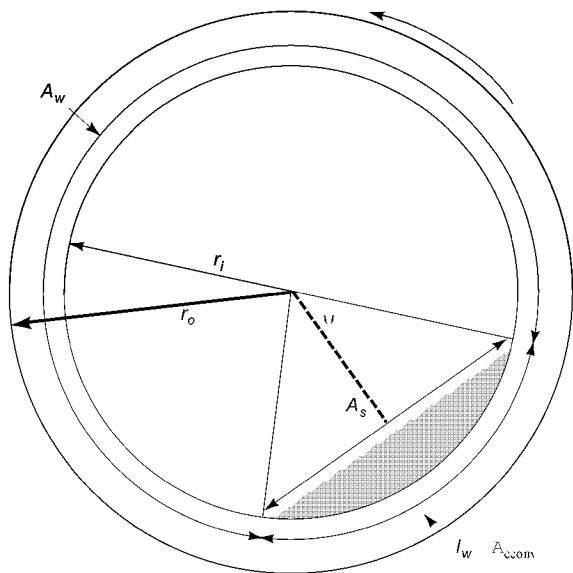


Fig. 4. Cross section of a rotary kiln. I_w is the wall area that is in contact with the solids (see eq. 10); A_w is the wall area that is in contact with the gases; A_s is the radiation sink area; r_i and r_o are the inner and outer radii, respectively. Other terms are defined in text.

Heat Transfer in Rotary Kilns. Heat transfer in rotary kilns occurs by conduction, convection, and radiation. In a highly simplified model, the treatment of radiation can be explained by applying a one-dimensional furnace approximation (19). The gas is assumed to be in plug flow; the absorptivity, α_g , and emissivity, ϵ_g , of the gas are assumed equal ($\alpha_g = \epsilon_g$); and the presence of water in the solids is taken into account. Energy balances are performed on both the gas and solid streams. Parallel or countercurrent kilns can be specified.

The solids are also assumed to be in plug flow. As part of the plug flow approximation, the gas and solids are assumed isothermal in the radial direction at a given axial location. Detailed models for kiln heat transfer are available (20,21).

A key part of the model is the energy balance on the solids which must account for the latent heat of vaporization of water in the bed. This balance assumes that the heating of the solids occurs in three stages: (1) the heating of the bed to 100°C; (2) a constant temperature period during water evaporation at 100°C; and (3) a final period during which the bed is completely dry. The first stage assumes that the rate of water loss is negligible until the bed temperature reaches 100°C. A number of heat-transfer coefficients are needed for the energy balances. The wall-to-solids heat-transfer coefficient, h_{ws} , is estimated from

$$\frac{h_{ws} I_w}{k_s} = 11.6 \left[\frac{n r^2 \beta}{a} \right]^{0.3} \tag{7}$$

where n is the kiln rotation rate in revolutions per second, β is twice ψ , a is the thermal diffusivity of the solids, r is the kiln's inner radius, k_s is the thermal conductivity of the solids, and I_w is defined in Figure 4. Equation 7 is valid for $n r^2 \beta / a < 10^4$ (22).

The gas-to-wall heat-transfer coefficient, h_{gw} , is estimated by

$$h_{gw} = 0.036 \frac{k_g}{d} Re_d^{0.8} Pr^{0.33} \left[\frac{d}{l} \right]^{0.055} \tag{8}$$

where l is the distance from the burner wall, d is the kiln's inner diameter, k_g is the thermal conductivity of the gas, Re is the Reynold's number, and Pr is the Prandtl number (23). Equation 8 is valid only for turbulent conditions, ie, $Re > 10,000$.

The gas-to-solids heat-transfer coefficient, h_{gs} , is the least certain:

$$h_{gs} = 0.4 (G'_g)^{0.62} \tag{9}$$

where the units on h_{gs} are $W / (m^2 \cdot K)$ and G'_g is the gas mass flux, $kg / (m^2 \cdot h)$ (23).

The time constants characterizing heat transfer in convection or radiation dominated rotary kilns are readily developed using less general heat-transfer models than that presented herein. These time constants define simple scaling laws which can be used to estimate the effects of fill fraction, kiln diameter, moisture, and rotation rate on the temperatures of the solids. Criteria can also be established for estimating the relative importance of radiation and convection. In the following analysis, the kiln wall temperature, T_w , and the kiln gas temperature, T_g , are considered constant. Separate analyses are conducted for dry and wet conditions.

If the dominant mode of heat transfer to the solids is convection between the wall and the solids, then the characteristic time for a dry system is

$$t_o = \frac{\rho c_{pds} V}{h_{ws} A_{conv}} \tag{10}$$

where the solids volume per unit length is $V = r^2 [\psi - \frac{1}{2} \sin(2\psi)]$, the area of the solids in contact with the wall, per unit length, is $A_{conv} = 2r\psi$, ρ is the solid density, and c_{pds} is the heat capacity of the dry solids. The kiln's geometry is defined in Figure 4 and the heat-transfer coefficient is given by equation 7. Hence, for a constant fill fraction (constant ψ), time should scale as $t_o \propto r^{1.4} n^{-0.3}$. It can also be shown that $t_o \propto f^{0.90}$, where f is the fill fraction.

In a radiation dominated kiln environment, with hot combustion gases and reradiating walls, the characteristic time is

$$t_o = \frac{\rho c_{pds} V R_{tot}}{\sigma (T_g^2 + T_i^2) (T_g + T_i)} \tag{11}$$

where R_{tot} is defined by equation 10 and T_i is the initial temperature of the solids.

If the kiln may be considered an enclosure bounding an isothermal gray gas of emissivity, ϵ_g , with two bounding surfaces consisting of reradiating walls of area A_w , and of bed solids (the radiation sink) of area A_s , then the expression for R_{tot} becomes (19)

$$R_{tot} = \frac{1}{\frac{\epsilon_s}{\epsilon_s A_s} + \frac{1}{\epsilon_g \left\{ A_s + \frac{A_w}{1 + \epsilon_g \left[F_{ws} (1 - \epsilon_g) \right] \right\}}}$$

(12)

where ϵ_s is the emissivity of the solids, $A_s = 2r\sin\psi$, $A_w = 2r(\pi - \psi)$, and the view factor, $F_{ws} = A_s/A_w$. For a given fill fraction, f , gas emissivity, ϵ_g ; and solid emissivity, ϵ_s , the time constant is directly proportional to the kiln's radius, $t_o \propto r$. This conclusion neglects the effect of r on ϵ_g .

Comparison of equations 10 and 11 shows that the factors VR_{tot} and VA_{conv} are similar in their functional dependencies on kiln fill fraction. The dependence of t_o on fill fraction is almost linear and to a good approximation, $t_o \propto f^{0.87}$.

An estimate of the relative importance of convection and radiation can be obtained from the ratio of the radiation-to-convection transfer rates. This dimensionless number reduces to

$$N_{rc} = \frac{\epsilon_s \sigma \sin(\psi) (T + T_\infty) (T^2 + T_\infty^2)}{h_{ws} \psi}$$

(13)

where h_{ws} is given by equation 7 and T is the solids' temperature. Equation 13 shows that the relative importance of radiation increases as the temperatures of the wall and solids increase.

If the rate of moisture vaporization is controlled by the rate of heat transfer to the wet solid, then for convection dominated heat transfer, h_{ws} , at the boiling point of water, the characteristic time is

$$t_o = \frac{m_o \Delta H_{vap}}{h_{ws} A_{conv} (T_\infty - T_{bp})} \quad b g n o (14)$$

where m_o is the initial mass of water, m is the mass of water at any time t , M is the ratio m/m_o , T_{bp} is the boiling point of water, and h_{ws} is assumed independent of the moisture content of the solids. For radiation dominated heat transfer,

$$t_o = \frac{m_o \Delta H_{vap} R_{tot}}{\sigma (T_g^4 - T_{bp}^4)}$$

(15)

Hence, according to equations 14 and 15, doubling the moisture content doubles the time required for vaporization, assuming that h_{ws} , $T_\infty - T_g$, and R_{tot} are constant.

Comparisons of the complete heat-transfer model with pilot-scale rotary kiln data are shown in Figure 5 (21) for moisture levels ranging from 0 to 20 wt %. The tremendous thermal impact of moisture is clearly visible in the leveling of temperature profiles at 100°C.

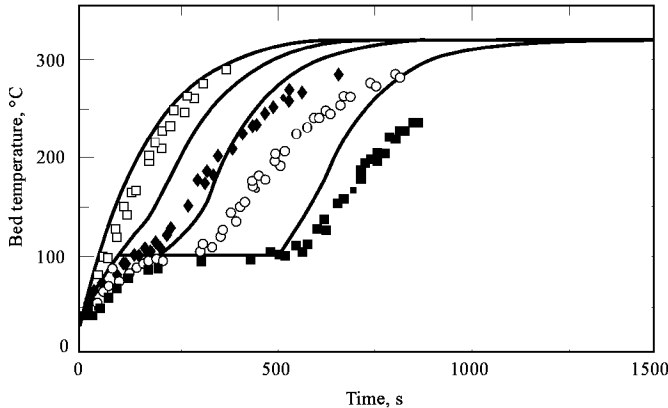


Fig. 5. The impact of moisture in the solids on the temperature profiles of the solids at a kiln wall temperature of 330°C, 0.5 rpm, and a fill fraction of 3%: (□) represents 0% H₂O; (◆), 5% H₂O; (○), 10% H₂O; and (■), 20% H₂O (21).

Mass Transfer and Kinetics in Rotary Kilns. The rates of mass transfer of gases and vapors to and from the solids in any thermal treatment process are critical to determining how long the waste must be treated. Oxygen must be transferred to the solids. However, mass transfer occurs in the context of a number of other processes as well. The complexity of the processes and the parallel nature of steps 2, 3, 4, and 5 of Figure 2, require that the parameters necessary for modeling the system be determined empirically. In this discussion the focus is on rotary kilns.

Free liquid hydrocarbons and water flash vaporize if they contact hot surfaces. A rough estimate of the magnitude of such an event can be made if the free liquid levels are known. Localized flashing of hydrocarbons and water continues, even in the absence of free liquids, whenever particles of waste are suddenly brought in contact with hot surfaces or exposed to intense radiation.

Aspects of the constant temperature period, step 3 of Figure 2, are described by equations 14 and 15, which give the time required to boil water from the bed. However, some or even all of the water and volatile hydrocarbons can leave the bed before the boiling point is reached throughout the bed.

The desorptive process may be analyzed before boiling. The key assumption is that the vapor and adsorbed phases are in equilibrium in the bulk of the bed. This assumption eliminates intraparticle resistances from further consideration and is reasonable for rotary kiln applications. The two remaining resistances are associated with hydrocarbon diffusion out of the bed and with convection from the bed surface to the bulk gases. The flux of species A from the desorbing bed becomes

$$N_A = k_B(C_A - C_{AS}) = k_F(C_{AS} - C_{A\infty})$$

(16)

where N_A = flux of A at the bed-freeboard interface, in mol (m²·s); k_B = bed-side mass-transfer coefficient, m/s; C_A = gas-phase concentration of A in bed's interior gases, mol/m³; C_{AS} = gas-phase concentration of A at bed-freeboard interface, mol/m³; k_F = freeboard-side mass-transfer coefficient, m/s; and $C_{A\infty}$ = concentration of A in bulk, freeboard gases, mol/m³. Eliminating C_{AS} from equation 16 and assuming that $C_{A\infty}$ is much less than C_{AS} gives

$$N_A = \frac{C_A}{\frac{1}{k_B} + \frac{1}{k_F}}$$

(17)

To estimate k_B , the slumping motion of the kiln bed which periodically exposes a fresh, vapor saturated surface at the bed–freeboard interface must be considered. Based on Fick's second law in a bed of porosity, ϵ , and for an effective diffusion coefficient, D_A , the mass-transfer coefficient on the bed side is

$$k_B = 2 \left(\frac{\epsilon D_A}{\pi \tau} \right)^{1/2}$$

(18)

where τ is the exposure time of the bed surface (15,18). Assuming a 3.7 × 11-m kiln having a 7% fill fraction, fired with methane at 8.8 × 10⁶ W at 50% excess air, it can be shown that k_F/k_B is about 100. Because k_F is so much greater than k_B , equation 17 becomes simply

$$N_A = k_B C_A$$

(19)

For pure benzene at 60°C the vapor pressure is about 53 kPa (7.7 psi), $C_A = 19 \text{ mol m}^{-3}$, and $N_A = 0.015 \text{ mol (m}^2\cdot\text{s)}^{-1}$ or about 1 g (m²·s)^{−1}. Equation 19 provides a rough estimate of actual fluxes and can be used to find the characteristic time for desorption. Let ω_A be the adsorbed phase concentration of A , mol/kg of clean dry solid. At low contamination levels, the vapor-phase concentration of A is related to ω_A by a linear isotherm (from eq. 2 at low p),

$$\omega_A = K C_A \text{ or } C_A = \omega_A / K$$

(20)

where K is a strong function of temperature (given by eq. 4) having units m³/kg. Performing a material balance on a thin section of the bed, perpendicular to the kiln axis as it moves down the kiln, gives the time scale for an isothermal system

$$t_o = \frac{K V_B \rho_B}{2 A_B} \left[\frac{\pi \tau}{\epsilon D_A} \right]^{1/2}$$

(21)

The ratio V_B/A_B is the average bed depth and can be calculated from

$$\frac{V_B}{A_B} = \frac{r^2 (\psi - 0.5 \sin 2\psi)}{2 r \sin \psi}$$

(22)

where r is the inside kiln radius and ψ is defined in Figure 4. ψ can be calculated from the fill fraction using

$$\psi = \pi f + 0.5 \sin 2\psi$$

(23)

The exposure interval for the bed, τ , is inversely proportional to the kiln rotation rate. Hence, equation 21 shows that the time constant for desorption is directly proportional to the bed depth and inversely proportional to the square root of the kiln rotation rate. However, the overriding factor affecting t_o is the isotherm constant K which in general decreases exponentially with increasing temperature as in equation 4.

Step 4 of the thermal treatment process (see Fig. 2) involves desorption, pyrolysis, and char formation. Much literature exists on the pyrolysis of coal (qv) and on different pyrolysis models for coal. These models are useful starting points for describing pyrolysis in kilns. For example, the devolatilization of coal is frequently modeled as competing chemical reactions (24). Another approach for modeling devolatilization uses a set of independent, first-order parallel reactions represented by a Gaussian distribution of activation energies (25).

The ability of a four-parameter, two-parallel reaction model to correlate pilot-scale rotary kiln, toluene-desorption results (26) is shown in Figure 6. The model assumes that the adsorbed toluene consists of two fractions, T and L , which are tightly and loosely bound, respectively.

$$\begin{aligned} \frac{dT}{dt} &= -k_T T \\ \frac{dL}{dt} &= -k_L L \end{aligned}$$

(24)

Each k is given by an Arrhenius expression, $k = A \exp(-E/RT)$, and the fraction of the tightly bound component is a parameter. For the high temperature results in Figure 6, some charring of toluene was observed at the highest wall temperature (790°C). The fraction of toluene remaining in the bed was determined from gas-phase total hydrocarbon, O₂, and CO₂ measurements.

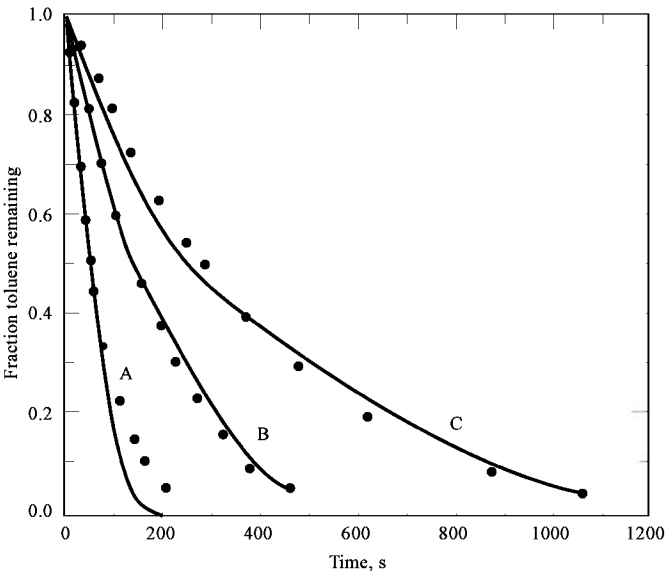


Fig. 6. Pilot-scale kiln results for a fill fraction of 0.08% at 0.5 rpm and an initial toluene loading, on a dry, calcined, montmorillonite clay adsorbent, of 0.25 wt %, at A, 790°C; B, 330°C; and C, 190°C. The solid lines are model fits using equation 24. The model simultaneously fits to all of the data (24).

The last step, step 5 (see Fig. 2), involves char oxidation. Oxygen transport to the kiln's solids occurs in two steps in series: convection of oxygen to the bed surface and diffusion of oxygen from the bed surface to the bed interior. Because of the slumping motion of the bed and the relatively slow diffusion of gases through the bed, the latter process can be described by a penetration model. From the prior analysis of the desorption of volatiles it is expected that the rate of mass transfer of oxygen is controlled by the rate of diffusion of oxygen into the bed. The mass-transfer coefficient is again given by equation 18 and calculations based on equation 18 show that char oxidation rates are exceptionally low, suggesting that stripping, desorption, and pyrolysis are the primary mechanisms for rendering the solids clean.

Data and modeling results show how bench-scale data and heat and mass-transfer models can be used to estimate full-scale performance. The example of hydrocarbon removal given in Figure 6 is for a dry clay. The addition of water is tremendously complicating. However, even in wet systems, provided that the kiln's wall temperatures are greater than ~400°C, distributed activation energy models or parallel reaction models still provide good correlations of desorption data and can be used for scaling purposes (15). Given the complexity of most solid wastes, desorption and or kinetic data are necessary for calibrating models of hydrocarbon removal and destruction.

Software for performing practical calculations related to hazardous waste incineration (HWI) is available (27), and performs three types of calculations: (1) thermochemical calculations such as material and energy balances relating to incinerator temperature, excess air, and feed heating value; (2) stoichiometric calculations to give exhaust gas compositions and flow rates; and (3) preliminary incinerator design calculations. Required input data for the design calculations include either the combustion gas velocity or the length-to-diameter ratio of the incinerator. The volumetric heat release rate for the facility is also required.

Pollutant Emissions from Solid Waste Incinerators.

NO_x and SO₂ Emissions. Oxides of nitrogen (NO and NO₂) and sulfur (primarily SO₂) are emitted from most combustion systems including hazardous waste incinerators. The two principle mechanisms is formed are summarized in Figure 7 (28). The thermal NO_x pathway is important in any high temperature process containing N₂ and O₂. The fuel NO pathway is also important if the fuel or waste contains nitrogen. NO_x levels are difficult to estimate because the rate of formation of NO by both pathways is determined by complex kinetics and gas-phase mixing. SO₂ production in incinerators is generally easy to predict because nearly all organic sulfur species are completely converted to SO₂ and other products of combustion. The incineration of liquid wastes containing nitrogen are fairly well understood from studies on liquid fuels (see COMBUSTION TECHNOLOGY).

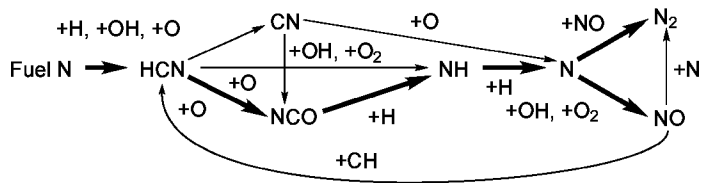


Fig. 7. Schematic showing reaction pathways by which fuel nitrogen, N, is converted to NO and N₂. The bold lines indicate the key pathways (28). Thermal NO is formed from N₂: N₂ + O → NO + N.

In a rotary kiln, the burner can produce both thermal and fuel NO_x, if the fuel contains nitrogen. Many solid waste streams also contain nitrogen, typically as much as 20 wt %, which contributes to the fuel NO pathway. Key sources of solid waste fuel nitrogen include plastics, nylons, dyes, and other process wastes. Nylon, for example, is 33 wt % nitrogen.

Batch, pilot-scale rotary kiln studies of NO_x formation have been performed by adding aniline [62-53-3], pyridine [110-86-1], ethylenediamine [107-15-3], or malononitrile [109-77-3] to an inert, calcined montmorillonite adsorbent (29). Initial studies were performed with aniline. The effect of the synthetic waste's nitrogen content on transient NO_x emissions from the 730°C kiln is shown in Figure 8. The tests were performed by adding toluene mixed with aniline to several kilograms of the adsorbent. The total weight of hydrocarbon added to the kiln in each experiment was fixed at 35 g. The actual nitrogen content of the solids was less than 0.1 wt %. Figure 8 shows that increasing the waste's nitrogen content increases NO_x levels significantly, but that the emission levels are not directly proportional to nitrogen content.

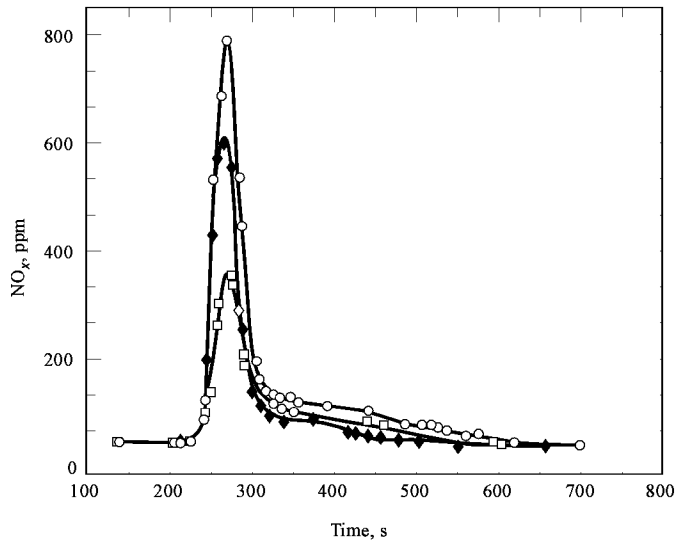


Fig. 8. Transient NO_x emissions at synthetic wastes' nitrogen contents of (□), 1%; (◆), 3%; and (○), 10%, where the % of O₂ is zero. Aniline is used as a nitrogen source and toluene is used to keep the total hydrocarbon weight fixed at 35 g (29).

In fact, the fractional conversion of the waste's nitrogen to NO_x decreases with increasing nitrogen content (see Fig. 8) (29), as can be understood from the reaction pathway (see Fig. 7).

$$N + NO \rightarrow N_2 + O \tag{25}$$

which converts NO to N₂. Increasing the NO concentration favors the destruction of N. Figure 9 shows that for a given nitrogen percentage, the source

of the nitrogen causes considerable variation in the NO_x levels. The variation results from the markedly different desorption time scales which characterize the different hydrocarbons. Hence, local stoichiometries are different for different nitrogen compounds, leading to changes in the NO_x chemistry. This is in contrast to liquid waste combustion in which the source compound for the nitrogen at a fixed weight percentage of nitrogen usually has little effect on NO_x levels.

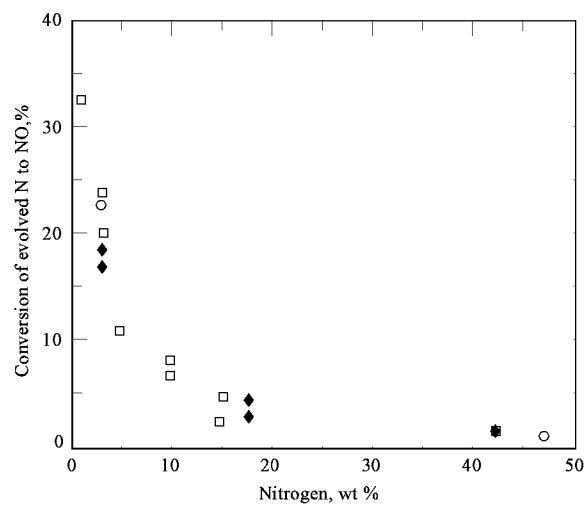


Fig. 9. Waste nitrogen conversion to NO as a function of the nitrogen content of the synthetic solid waste where (□) represents aniline; (◆), pyridine; (○), ethylenediamine; and (■), malononitrile (29).

Partitioning of Heavy Metals. Metals entering a solid waste incinerator can leave the system with the bottom ash, the captured fly ash, or the exhaust gases. The fraction leaving with the exhaust gases can include metal vapors such as mercury (qv) and submicrometer particles that escape capture in the air pollution control devices. Metals entering the incinerator as liquid streams are usually carried out of the reactor in the fly ash, which is typically enriched with heavy metals relative to the entering solid waste. The level of enrichment increases with increasing incinerator temperature and increasing metal volatility. Pilot-scale rotary kiln data showing both effects are given in Figure 10 (30). The enrichment on small particles results from the condensation of vaporized metals. Metal concentrations on particles also typically increase with decreasing particle diameter. The vast literature on the behavior of metals in coal combustion is useful for understanding these phenomena (31). A comprehensive review on metal emissions from incinerators is available (32).

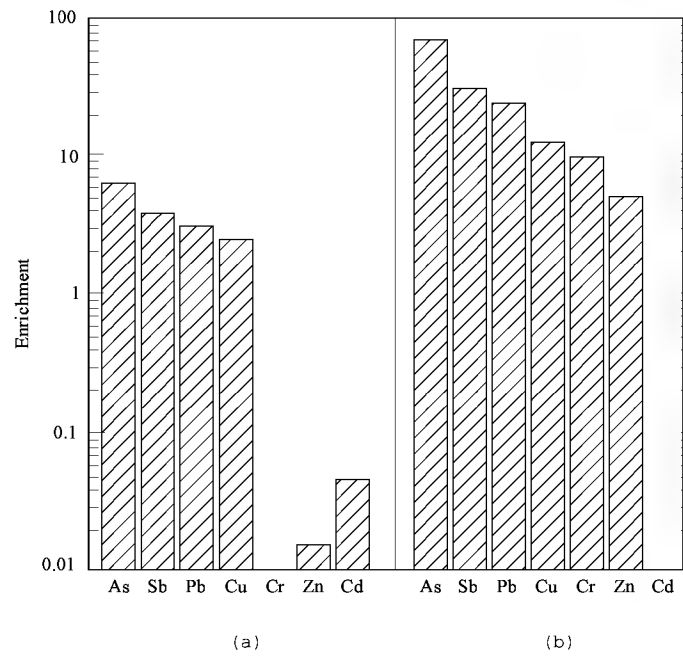


Fig. 10. The enrichment of metals in the fly ash of a pilot-scale rotary kiln as a function of temperature (a) at 540°C and (b) at 980°C (30).

Dioxin and Furan Emissions. The emissions of polychlorinated dibenzo-*p*-dioxins (PCDD) and polychlorinated dibenzo-furans (PCDF) from incinerators (4) are of interest to the public, scientists, and engineers. The U.S. EPA classifies 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (2,3,7,8-TCDD) as the most potent carcinogenic compound it has evaluated. It is also listed as the agency's most potent reproductive toxin (4).

The proposed mechanism by which chlorinated dioxins and furans form has shifted from one of incomplete destruction of the waste to one of low temperature, downstream formation on fly ash particles (33). Two mechanisms are proposed, a *de novo* synthesis, in which PCDD and PCDF are formed from organic carbon sources and Cl in the presence of metal catalysts, and a more direct synthesis from chlorinated organic precursors, again involving heterogeneous catalysis. Bench-scale tests suggest that the optimum temperature for PCDD and PCDF formation in the presence of fly ash is roughly 300°C.

Chlorine may be formed by the Deacon reaction at temperatures below about 900°C,



Both Cl₂ and HCl have been shown to chlorinate hydrocarbons on fly ash particles. Pilot-scale data involving the injection of fly ash from municipal waste combustion (33) show that intermediate oxygen concentrations (4–7%) produce the highest levels of PCDD and PCDF. These data also show significant reductions in PCDD and PCDF emissions with the upstream injection of Ca(OH)₂ at about 800°C.

Liquid Waste Incineration

Incinerators.

Furnaces. A furnace for combusting both high and low heating value liquid wastes is shown in Figure 11. Vertical furnaces are normally used for wastes containing high salt concentrations. Investment is typically higher than for furnaces of horizontal orientation as burners and controls are located in an elevated position, installation of furnace refractory is more difficult, and additional structural steel to support the furnace is required. For systems having a quench tank downstream of the furnace, the outlet of the furnace is tapered to reduce exposure to the quench and subsequent radiation losses. Furnace outlet velocities are maintained below 24 m/s to minimize erosion of outlet refractory.

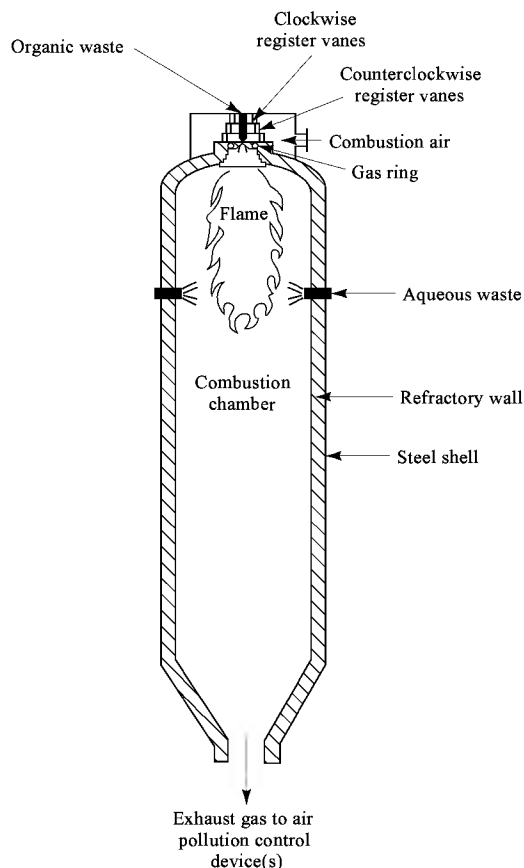


Fig. 11. Liquid waste incinerator furnace.

For given combustion air, waste, and auxiliary fuel feed rates to the incinerator, furnace residence time decreases as furnace pressure decreases. Often the required pressure drop through the downstream particulate removal device is not established until actual operation, and furnaces are sized assuming little or no pressure drop across the particulate removal device to allow adequate furnace residence time under all operating conditions.

Regulations require that the incinerator furnace be at normal operating conditions, including furnace temperature, before hazardous wastes are injected. This requires auxiliary fuel burners for furnace preheating. In addition, the burners provide heat when the wastes burned are of low heating value. Auxiliary burners are sized for conditions where liquid wastes are injected without the addition of high heating value wastes.

Auxiliary fuels normally include natural gas and No. 2 or No. 6 fuel oils. Natural gas has several advantages: it is relatively clean burning and easily transported and measured. No. 2 fuel oil requires mechanical, steam, or air atomization to ensure combustion. Moreover, No. 2 fuel oil burners, when not properly adjusted, tend to form soot and increase formation of products of incomplete combustion (PICs). No. 6 fuel oil is highly viscous, normally in the 400–2000 mPa·s (cP) range, at ambient temperatures, and as a result requires preheating to 80–120°C to reduce the viscosity to 15–65 mPa·s (cP) prior to combustion. No. 6 fuel oil piping requires steam or electric tracing to ensure lower viscosities as the fuel is being injected into the incinerator furnace. It normally requires burners capable of steam atomization. In addition, it can contain up to 3% sulfur and 500 ppm vanadium, which adds to the sulfur dioxide and vanadium pentoxide loading, and corrosiveness of flue gases leaving the combustion chamber.

The furnace is constructed with a steel shell lined with high temperature refractory (see REFRATORIES). Refractory type and thickness are determined by the particular need. Where combustion products include corrosive gases such as sulfur dioxide or hydrogen chloride, furnace shell temperatures are maintained above about 150–180°C to prevent condensation and corrosion on the inside carbon steel surfaces. Where corrosive gases are not present, insulation is sized to maintain a shell temperature below 60°C to protect personnel.

Three types of refractory are used. Castable refractory, similar to concrete, is placed in the shell using forms and poured in place or blown in. Plastic refractory is prepared in a stiff consistency and is either hammered or rammed in place. Plastic refractories are typically used for repairs. Fire brick is the most commonly used refractory. It is bonded in place using thin mortar joints. Brick having alumina content of 60–90% is used in areas exposed to hot corrosive gases. To decrease investment, less expensive insulating brick is often placed between the high alumina brick and the furnace shell. The cost of high alumina brick is typically 4–6 times that of insulating brick.

Refractory failures resulting from erosion and corrosion from hot particulate laden gases can result in incinerator downtime and high maintenance costs. Of particular concern are fluorine, sodium, potassium, and sulfate salts, which penetrate brick surfaces when hot. Upon cooldown, salt hardens and expands, causing the surface which has been penetrated to fail. In addition, organically bound alkali metals in wastes can react chemically with the refractory to form new compounds with lower melt points (eutectics) than furnace operating temperatures. Continued operation at elevated furnace temperatures and close attention to the design and operation of the furnace to keep wastes from impinging on refractory walls, along with controlling the amount of alkali metals fed, help prolong refractory life.

Quench Systems. Quench systems are used to cool hot furnace gases from 980–1200°C to 120–150°C. This allows less expensive materials of construction such as fiber glass reinforced plastic (FRP) (see GLASSES, ORGANIC INORGANIC HYBRIDS) to be used downstream of the quench and in gas cleaning equipment, and reduces the volume of gas flow, resulting in smaller equipment. Water or air quenching systems are typically used. Water quenching systems use the latent heat of evaporation to adiabatically cool gases. Particulate matter collects in quench water requiring that the system be continuously purged. Air quench systems require the addition of large volumes of ambient air resulting in larger downstream gas cleaning equipment than with water quench systems. Air quench chambers are sized to allow particulate matter to be removed manually during incinerator shutdown. Materials of construction and design of spray systems are critical in the design of quench systems.

Control Systems. Control systems are used to regulate the addition of liquid waste feed, auxiliary fuel, and combustion air flows to the incinerator furnace. In addition, scrubber operation is automated to help ensure meeting emission limits. Flows are measured using differential pressure

devices such as flow nozzles and flow orifices when the fluids are relatively nonviscous ($Re > 3000$). Gas and combustion air flows are typically measured using a flow orifice. Highly viscous wastes require measurement with positive displacement systems such as oval gear, target flow, or mass flow meters (see FLOW MEASUREMENTS).

Temperature measurements ranging from 760 to 1760°C are made using iron–constantan or chromel–alumel thermocouples and optical or surface pyrometers. Temperature measuring devices are placed in multiple locations and protected to allow replacement without incinerator shutdown (see TEMPERATURE MEASUREMENT).

To ensure combustion gases are not present during initial burner light off, the furnace is purged with ambient air. After the purge, a 145–585 kJ s (0.5–2.0 MBtu h) pilot establishes a flame in the furnace. A scanner is used to sense flame and is interlocked to shut down all waste and auxiliary fuel flows upon loss of flame. Flame scanners for incinerator furnaces are typically of the ultraviolet type because hot refractory emits in the infrared region. After the pilot has been established, the auxiliary fuel burners are ignited and the furnace is brought up to its normal operating temperature. Once the operating conditions given in the permit are satisfied, including furnace temperature, stack oxygen levels, and combustion air flow rates, wastes are injected into the furnace.

Factors Affecting Destruction of Liquid Wastes. Liquid wastes can be divided into two classes: low and high heating value. The former requires auxiliary fuel. A heating value above 16 MJ/kg (7000 Btu/lb) is generally considered high enough to be burned without auxiliary fuel, but this depends on the specifics of excess air and desired flame temperature. As shown in Figure 11, high heating value wastes are usually injected with auxiliary fuel and oxidant into a swirl burner. Roughly one vessel diameter is allowed in the axial direction for mixing and combustion of the fuels, and then the low heating value waste, which is typically aqueous, is injected radially into the hot gases. Because the high heating value waste is injected with the fuel, more residence time is available for its destruction, so that the destruction of the low heating value waste is generally limiting. Similar mechanisms are involved for both types of waste.

The steps in waste destruction are (1) heatup of the waste to its boiling point, (2) vaporization of the waste droplets, (3) heatup of any waste residue to combustion temperature, and (4) destruction by combustion reaction. Destruction of waste species volatilized during the first three steps can begin the moment a species becomes gaseous, because sensible heating of volatilized components is extremely rapid, and need not be considered as a separate step.

The factors which govern the efficiency of waste destruction include atomization, ie, mean drop size, and size distribution; temperature; residence time; O₂ concentration; and flow patterns.

The atomization, temperature, and O₂ concentration interact with the various rate processes to set the residence time requirements. A desirable flow pattern is for the fuel, ie, the high heating value waste plus any supplemental fuel, to burn completely in a backmixed zone, and for the low heating value waste to be injected into a plug flow zone having uniform temperature and O₂ concentration profiles. This pattern generally produces the minimum average residence time requirements. To the extent that part of the incinerator volume is ineffective owing to cold spots, O₂ starved spots, or bypassing of portions of the flow, the average residence time has to be increased to compensate. In cases where the waste contains dissolved or suspended solids, a design allowance should be made for the probability that these solids slowly coat the chamber walls and reduce the effective residence time during the course of operation. Periodic burnouts or shutdowns for cleaning are required to remove this material.

Droplet Heatup. A relation for the time required for droplet heatup, τ_h , can be derived based on the assumption that forced convection is the primary heat-transfer mechanism, and that the Ranz-Marshall equation for heat transfer to submerged spheres holds (34). The result is

$$\tau_h = \frac{d_o^2}{k_h} \tag{27}$$

in which d_o is the drop diameter and k_h is the heatup rate constant, calculated from

$$k_h = \frac{12 \lambda_g (1 + \phi)}{\rho_l c_{pl} \ln \left[\left(\frac{T_g - T_{d,0}}{T_g - T_{d,b}} \right) \right]} \tag{28}$$

In this equation, λ_g is the gas thermal conductivity; ρ_l the liquid density; c_{pl} the liquid heat capacity; T_g the gas temperature; $T_{d,0}$ the initial droplet temperature; and $T_{d,b}$ the droplet boiling point.

To the extent that radiation contributes to droplet heatup, equation 28 gives a conservative estimate of the time requirements. The parameter ϕ reflects the dependence of the convective heat-transfer coefficient on the Reynolds number:

$$\phi = 0.3 (Re_o)^{\frac{1}{2}} (Pr_g)^{\frac{1}{3}} \tag{29}$$

where Pr_g is the gas-phase Prandtl number, and Re_o the Reynolds number for the initial drop size. This latter is calculated at its terminal velocity in a gravitational field assuming laminar flow past the drops:

$$Re_o = \frac{d_o^3 (\rho_l - \rho_g) \rho_g g}{18 \mu_g^2} \tag{30}$$

where ρ_g is the gas density, μ_g is the gas viscosity, and g is the gravitational acceleration.

Droplet Evaporation. Based on the same assumptions, the time for droplet evaporation, τ_e , can be estimated as

$$\tau_e = \frac{d_o^2}{k_e} \tag{31}$$

The evaporation rate constant, k_e , is estimated via

$$k_e = \frac{8 \lambda_g (T_g - T_{d,b}) f(\phi)}{\rho_l \Delta H_{vl}} \tag{32}$$

where ΔH_{vl} is the heat of vaporization of the droplet liquid.

The function $f(\phi)$ arises because the droplet is shrinking, and the Reynolds number changes along the evaporation trajectory. In the course of integrating the underlying differential equation, two limiting cases can be solved analytically giving $f(\phi) = \phi^4$ for $\phi \gg 1$ and $f(\phi) = 1$ for $\phi < 1$. The function $f(\phi)$ allows interpolation between these two limiting values to get an average k_e along the evaporation trajectory:

$$f(\phi) = \frac{1 + 0.995\phi + 0.174\phi^2 + 0.00113\phi^3}{1 + 0.423\phi + 0.00453\phi^2} \tag{33}$$

Residue Heatup. Equations 27–30 can be used to estimate the time for residue heatup, τ_{th} , by replacing the liquid properties, such as density and heat capacity, with residue properties, and considering the now smaller particle in evaluating the expressions for Re_ϕ , ϕ , k_ϕ , and τ_ϕ . In the denominator of k_ϕ , $T_{d,0}$ is replaced by $T_{d,b}$ and $T_{d,b}$ is replaced by $T_{r,\phi}$ the ignition temperature of the residue.

Waste Destruction by Combustion Reaction. At the simplest level, waste destruction can be thought of as a first-order process involving the thermally excited rupture of a chemical bond. The time to achieve a given extent of reaction, τ_p , can be found from

$$\tau_r = \frac{\ln(1 - \epsilon)}{k_r} \tag{34}$$

in which ϵ is the fraction converted and k_r is the reaction rate constant. The rate constant is usually represented in the familiar Arrhenius form:

$$k_r = k_{r,0} e^{-E/RT} \tag{35}$$

For true first-order bond rupture reactions, the activation energy, E , is equal to the energy of the ruptured bond, and following the transition-state theory (35), the pre-exponential factor, $k_{r,0}$, is

$$k_{r,0} = \frac{kT}{h} \left(e^{S_a/R} \right) \tag{36}$$

in which k is the Boltzmann constant, h is the Planck constant, S_a is the entropy change of activation, and R is the universal gas constant. Taking an average for simple fission reactions gives $S_a/R = 6$ (36), and $k_{r,0}$ becomes $1.1 \times 10^{16} \text{ s}^{-1}$ at 980°C.

The characteristic times for waste destruction to an efficiency of 99.99%, for water in a nitrogen atmosphere, where the residue is a typical hydrocarbon breaking into two large fragments, eg, *n*-butane decomposing into two ethyl radicals would be droplet heatup, 0.073 s; droplet evaporation, 1.0 s; residue heatup, 0.071 s; and reaction, 0.23 s to give a total time of 1.37 s. The initial aqueous droplet is assumed to be 400 μm in diameter leaving a 100-μm diameter residue upon evaporation. The activation energy of the decomposition is assumed to be ca 350 kJ/mol (83 kcal/mol).

The overall requirement is 1.0–2.0 s for low energy waste compared to typical design standards of 2.0 s for RCRA hazardous waste units. The most important, ie, rate limiting steps are droplet evaporation and chemical reaction. The calculated time requirements for these steps are only approximations and subject to error. For example, formation of a skin on the evaporating droplet may inhibit evaporation compared to the theory, whereas secondary atomization may accelerate it. Errors in estimates of the activation energy can significantly alter the chemical reaction rate constant, and the pre-exponential factor from equation 36 is only approximate. Also, interactions with free-radical species may accelerate the rate of chemical reaction over that estimated solely as a result of thermal excitation; therefore, measurements of the time requirements are desirable.

Droplet Evaporation Measurements. Droplet evaporation rate measurements can be made by suspending a waste drop of known initial size from a thermocouple junction over a flame flowing at known velocity as shown in Figure 12. A video tape of the evaporation process can be used to track drop size with time. A plot of the square of the diameter with time gives a straight line with a negative slope equal to the evaporation rate constant as shown in Figure 13. The droplet boiling point is recorded as the drop evaporates and the final temperature after complete evaporation gives an indication of gas temperature. Such an experiment also yields an estimate of the residue diameter as a fraction of the initial drop diameter.

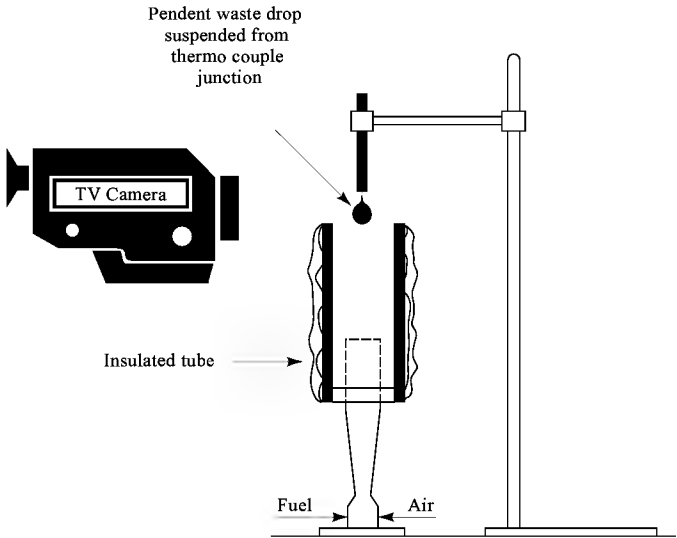


Fig. 12. Droplet evaporation measurement apparatus.

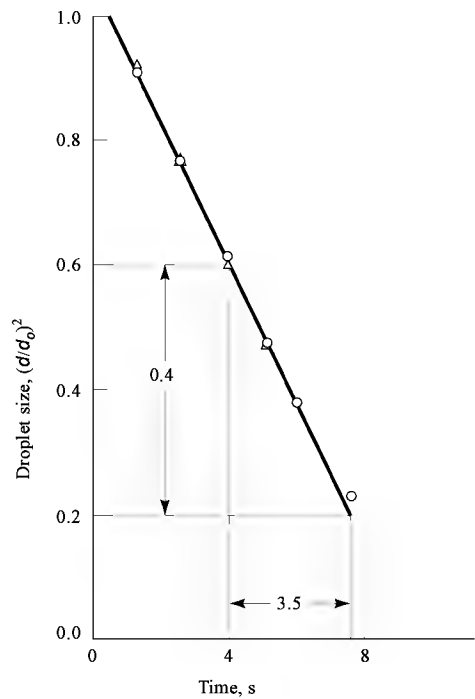


Fig. 13. Droplet evaporation rate constant where $d_0 = 0.18\text{ cm}$ and k_e can be shown to be $0.4(0.18)^2 \cdot 3.5 = 0.0037\text{ cm}^2\text{ s}$ (see eq. 31).

Chemical Reaction Measurements. Experimental studies of incineration kinetics have been described (37–39), where the waste species is generally introduced as a gas in a large excess of oxidant so that the oxidant concentration is constant, and the heat of reaction is negligible compared to the heat flux required to maintain the reacting mixture at temperature. The reaction is conducted in an externally heated reactor so that the temperature can be controlled to a known value and both oxidant concentration and temperature can be easily varied. The experimental reactor is generally a long tube of small diameter so that the residence time is well defined and axial dispersion may be neglected as a source of variation. Off-gas analysis is used to track both the disappearance of the feed material and the appearance and disappearance of any products of incomplete combustion.

The classical experiment tracks the off-gas composition as a function of temperature at fixed residence time and oxidant level. Treating feed disappearance as first order, the pre-exponential factor (k_e) and activation energy, E , in the Arrhenius expression (eq. 35) can be obtained. These studies tend to confirm large activation energies typical of the bond rupture mechanism assumed earlier. However, an accelerating effect of the oxidant is also evident in some results, so that the thermal rupture mechanism probably overestimates the time requirement by as much as several orders of magnitude (39). Measurements at several levels of oxidant concentration are useful for determining how important it is to maintain spatial uniformity of oxidant concentration in the incinerator.

Reaction measurement studies also show that the chemistry is often not a simple one-step reaction process (37). There are usually several key intermediates, and the reaction is better thought of as a network of series and parallel steps. Kinetic parameters for each of the steps can be derived from the data. The appearance of these intermediates can add to the time required to achieve a desired level of total breakdown to the simple, thermodynamically stable products, eg, CO_2 , H_2O , or N_2 .

Atomization. Droplet heatup and evaporation calculations can be done for any droplet size, but are most often carried out to reflect the behavior of a mean-sized droplet. The finer the droplet, the less time required for the various steps in the destruction of the waste.

There are several practical problems related to atomization. First, if the droplet size distribution is broad, the largest droplets can contain a significant fraction of the waste, and can take much longer to heat up and evaporate than the mean-sized drop. Thus knowledge of the size distribution produced by the injection atomizer is important, and time estimates should be based on the larger droplets. Droplets that are too fine can evaporate too quickly, ie, before the droplet cloud has mixed well with the oxidant-bearing gas. This condition, which can lead to reaction under starved conditions and lead to high potential for sooting, should be avoided. A narrow droplet size distribution is best.

The spray pattern is also important. A solid cone pattern mixes less well than a hollow cone, sheet, or multiple jet patterns. Drops that enter along the edge of the spray pattern nearest the exit have less time to heat up and evaporate. Drops that enter along the edges of the spray pattern nearest the walls may not fully evaporate before hitting the walls, resulting in erosion and corrosion of the brick.

Droplet trajectories for limiting cases can be calculated by combining the equations of motion with the droplet evaporation rate equation to assess the likelihood that drops exit or hit the wall before evaporating. It is best to consider upper bound droplet sizes in addition to the mean size in these calculations. If desired, an instantaneous value for the evaporation rate constant may also be used based on an instantaneous Reynolds number calculated not from the terminal velocity but at a resultant velocity. In this case, equation 37 is substituted for equation 32:

$$k_e = \frac{8\lambda_g(T_g - T_{d,b}) \left(1 + 0.3Re^{1/2}Pr_g^{1/3}\right)}{\rho_l \Delta H_{vl}} \tag{37}$$

Computer Models. The actual residence time for waste destruction can be quite different from the superficial value calculated by dividing the chamber volume by the volumetric flow rate. The large activation energies for chemical reaction, and the sensitivity of reaction rates to oxidant concentration, mean that the presence of cold spots or oxidant deficient zones render such subvolumes ineffective. Poor flow patterns, ie, dead zones and bypassing, can also contribute to loss of effective volume. The tools of computational fluid dynamics (qv) are useful in assessing the extent to which the actual profiles of velocity, temperature, and oxidant concentration deviate from the ideal (40).

Several design strategies are available. The use of swirl burners for the supplemental fuel and high heating value waste can promote backmixing near the burner and prevent core bypassing. Too much swirl is to be avoided because it induces bypassing up the walls. Too little swirl results in a long, narrow flame with dead zones near the walls. An optimum range exists (41). The use of swirl is aimed at maximizing the effective volume fraction. A second approach is to estimate the effective volume fraction from the modeling, and to increase the total incinerator volume to achieve the desired effective volume.

Pollutant Emissions from Liquid Waste Incinerators.
Gas Emissions. Wastes considered for incineration are usually organic in nature, so that the vast majority of the waste ends up as CO_2 , H_2O , and N_2 . Although CO_2 is of increasing concern because it is a greenhouse gas (see *ATMOSPHERIC MODELS*), emission issues have generally been related to the combustion products of sulfur, halogens, and metallic components of the waste. Carbon monoxide, which is not thermodynamically stable under normal incineration conditions, is also regulated. If CO is found in incinerator off-gas it is evidence of poor fuel–oxidant mixing or insufficient effective residence time, possibly owing to cold spots, oxidant starved zones, or bypassing.

Sulfur generally becomes SO₂, although some smaller amounts are possibly converted to SO₃, depending on temperature. Chlorine mostly results in HCl, but some Cl₂ and atomic Cl forms as well. Any atomic Cl recombines to form Cl₂ if quenching is rapid. Low incineration temperatures favor Cl₂, and high temperatures favor atomic Cl. There is an optimal temperature for minimizing the total effective Cl₂, ie, Cl₂ + Cl₂.

HCl can be absorbed into water to make a concentrated (usually 20 wt %) HCl solution, whereas SO₂ and Cl₂ must be scrubbed using a basic reagent (usually caustic or lime) to be effectively removed. The scrub liquor pH must be controlled to avoid heavy reagent consumption caused by the removal of CO₂. Even if only HCl is present, it is usually a good idea to follow the HCl absorber with a basic scrubber in order to ensure complete elimination of acids from the exhaust gas. When absorbing chlorine into alkaline solutions, the products are both chloride and hypochlorite. It is necessary to keep the pH above 9 in order to stabilize the hypochlorite and avoid regeneration of chlorine. The blowdown can be treated in one of several ways. Hypochlorite can be decomposed to chloride and oxygen using homogenous transition-metal catalysts (42), reduced to chloride and sulfate with sulfite, or reacted with hydrogen peroxide to form chloride, water, and oxygen.

Oxides of nitrogen, NO_x, can also form. These are generally at low levels and too low an oxidation state to consider water scrubbing. A basic reagent picks up the NO₂, but not the lower oxidation states; the principal oxide is usually NO, not NO₂. Generally, control of NO_x is achieved by control of the combustion process to minimize NO_x, ie, avoidance of high temperatures in combination with high oxidant concentrations, and if abatement is required, various approaches specific to NO_x have been employed. Examples are NH₃ injection and catalytic abatement (43).

Engineering Calculations. A good discussion of column design for gas absorption (qv) can be found in the literature (44). For staged columns (tray or plate) the classic approach is to estimate the number of equilibrium stages, and translate this to the number of actual stages through models for stage efficiencies. For packed towers, either the height of a theoretical plate is used to translate an equilibrium-stage calculation to a packing height, or the concept of the transfer unit is used to estimate the number or required transfer units, and the height of a transfer unit.

These design methods depend on a knowledge of the vapor–liquid equilibria for estimating the number of equilibrium stages or the number of transfer units. Usually, the scrub liquor is an aqueous electrolyte, possibly containing suspended solids. Techniques for predicting the pertinent equilibria have improved to the point where reliable calculations can be done on solutions of the high ionic strength typical of these scrub liquors (45). The techniques are available in computer models of countercurrent contactors so that the number of required equilibrium stages can be readily estimated (see also ENGINEERING, CHEMICAL DATA CORRELATION) (46).

Metal Contaminants and Ash. Alkali metals form basic oxides that are very reactive toward acidic species such as the acid gases, silicates, and aluminates. These form stable salts with acid gases if the off-gas contains such gases. Sodium, the most common of these metals, prefers to form chlorides ahead of sulfates. Sodium carbonate only forms in the absence of halides and sulfur oxides, SO_x. There usually is too little NO_x present to form nitrates (see SODIUM COMPOUNDS).

Alkali metal halides can be volatile at incineration temperatures. Rapid quenching of volatile salts results in the formation of a submicrometer aerosol which must be removed or else exhaust stack opacity is likely to exceed allowed limits. Sulfates have low volatility and should end up in the ash. Alkaline earths also form basic oxides. Calcium is the most common and sulfates are formed ahead of halides. Calcium carbonate is not stable at incineration temperatures (see CALCIUM COMPOUNDS). Transition metals are more likely to form an oxide ash. Iron (qv), for example, forms ferric oxide in preference to halides, sulfates, or carbonates. Silica and alumina form complexes with the basic oxides, eg, alkali metals, alkaline earths, and some transition-metal oxidation states, in the ash.

Estimates of Composition. The best approach toward estimating the chemistry of most contaminant species is to assume chemical equilibrium. Computer programs and databases (qv) for calculating chemical equilibria are widely available (47). Care must be taken that all species of concern are in the database referenced by the program being used, and if necessary, important species must be added in order to get the complete picture. In addition to predicting the exhaust composition of both gases and solids, the ability of these chemical equilibrium programs to do adiabatic calculations makes them useful for computing supplemental fuel requirements and the effect of excess oxidant on temperature.

The equilibrium approach should not be used for species that are highly sensitive to variations in residence time, oxidant concentration, or temperature, or for species which clearly do not reach equilibrium. There are at least three classes of compounds that cannot be estimated well by assuming equilibrium: CO, products of incomplete combustion (PICs), and NO_x. Under most incineration conditions, chemical equilibrium results in virtually no CO or PICs, as required by regulations. Thus success depends on achieving a nearly complete approach to equilibrium. Calculations depend on detailed knowledge of the reaction network, its kinetics, the mixing patterns, and the temperature, oxidant, and velocity profiles.

NO_x formation occurs by a complex reaction network of over 100 free-radical reactions, and is highly dependent on the form of nitrogen in the waste. Nitro-compounds form NO₂ first, and then NO, approaching equilibrium from the oxidized side. Amines form cyano intermediates on their way to NO, approaching equilibrium from the reduced side. Using air as the oxidant, NO_x also forms from N₂ and O₂. This last is known as thermal NO_x.

Through numerical codes which integrate stiff differential equations (48), it is possible to estimate NO_x levels based on kinetic schemes. The reaction network is becoming better defined, both in terms of the important reactions and their kinetic parameters. These estimates only show trends; comparisons with data available for compounds similar to those being incinerated are recommended. Because the reaction trajectory is sensitive to the type of nitrogen-bearing species, chemical similarity is the important factor in making good projections from literature data.

Particulate Pollutant Control Equipment.

Venturi Scrubbers. Venturi scrubbers consist of a convergent section, a throat, and a divergent section. Particulate laden gases enter the convergent section, accelerate to approximately 130 m/s (425 ft/s), and are mixed with water via a spray system at the throat. Smaller (<0.1 μm in dia) particles are removed by diffusion; larger particles agglomerate in the water mist. Gases flow to the divergent section where velocities decrease allowing agglomerated particulates to drop out. Approximately 50–70% of particulate matter leaving liquid waste incinerator furnaces is less than one micrometer in diameter, making removal difficult. To effect higher removal efficiencies, pressure drop between the Venturi inlet and throat is increased to promote greater turbulence and agglomeration. Typical pressure drops range from 2.5 kPa to more than 25 kPa (10–100 in. water). Removal efficiencies are in the 50–99% range depending on particulate size distribution and throat pressure drop. The system can be automated to raise or lower the pressure drop as required removal efficiencies change.

Electrostatic Precipitators. Both wet and dry electrostatic precipitators are generally used for polishing or small particulate removal. Wet precipitators consist of an ionizer section to impart an electrical charge on particulate matter, followed by a packed-bed or plate-wetted scrubber. Using these devices, particulate matter greater than about 3 μm in diameter is removed by impaction on the wetted bed, whereas electrically charged particulate material smaller than 3 μm in diameter is removed by electrostatic attraction to the bed material or water droplets. Pressure drops are typically less than 2.5 kPa (10 in. water). Dry electrostatic precipitators work on essentially the same principle; a positive electrical charge is imparted to particulate and then collected on charged plates. Collection efficiencies are a function of the uniformity of gas flows, both with time and across the field, the electrostatic field strength, and particulate matter electrical resistivity.

Baghouses. Baghouses include a series of bags suspended in a vertical orientation through which particulate-laden gases pass. Particulates are filtered out at the bag surface. The bags are periodically shaken or rapped to allow the dust collected to fall into a hopper below. Baghouses have potential for pluggage with salts or unburned wastes, and are temperature limited by bag materials. Materials are available that are acid resistant and withstand temperatures up to 260°C. In most cases, baghouses require a quench system upstream to cool furnace gases. Particulate removal efficiencies are in the 95–99% range.

Incinerators for Vapors and Gases

Undesirable combustible gases and vapors can be destroyed by heating to the autoignition temperature in the presence of sufficient oxygen to ensure complete oxidation to CO₂ and H₂O. Gas incinerators are applied to streams that are high energy, eg, pentane, or are too dilute to support combustion by themselves. The gas composition is limited typically to 25% or less of the lower explosive limit. Gases that are sufficiently concentrated to support

combustion are sometimes burned in flares, waste-heat boilers, in conjunction with other fuels in boilers and kilns, or used as process fuel. Occasionally, such gases may be burned in specially designed furnaces incorporating heat or material recovery, eg, chlorine. Protection against flame flashback must be provided between the incinerator and the process or source of the waste gas. Such protection is necessary to guard against unusual gas compositions even when the waste gas stream does not support combustion. Design of the flashback-prevention device depends on stream composition. Parallel-plate or multiple-screen flame arrestors, or a water seal, are usually adequate, but special precautions must be taken when the gas contains appreciable quantities of hydrogen because of high flame speed. A carrier gas such as nitrogen can be used to provide sufficient velocity to prevent flashback. Knockout devices should be considered to prevent carrying slugs of combustible entrained liquid into the incinerator. Safety controls are desirable to protect against overheating caused by erratic flows and sudden composition changes. Noncombustible particles can present problems. These may be melted and collected as a usually corrosive liquid pool. Removal of particulate prior to incineration may be desirable, but if the material is too fine, especially submicrometer in size, efficient removal may not be economical. In such cases, direct-flame incinerators specifically designed for slag removal may be preferred.

Catalytic Incinerators. Catalytic incinerators, often used to remove hydrocarbons from exhaust gas streams, are more compact than direct-flame incinerators, operate at lower temperatures, often require little fuel, and produce little or no NO_x from atmospheric fixation. However, the catalytic bed must be preheated and carefully temperature controlled. Thus these are generally unsuited to intermittent and highly variable gas flows.

Direct-Flame Incinerators. In direct-flame incineration, the waste gases are heated in a fuel-fired refractory-lined chamber to the autoignition temperature where oxidation occurs with or without a visible flame. A fuel flame aids mixing and ignition. Excess oxygen is required, because incomplete oxidation produces aldehydes, organic acids, carbon monoxide, carbon soot, and other undesirable materials.

A simple direct-flame incinerator may be a refractory-lined furnace arranged for good mixing and fitted with a burner. Such an incinerator is low in capital cost and suitable for periodic or batch burning of a process purge gas during plant shutdown. However, such an incinerator in continuous service on other than a very small vent would have extremely high fuel operating cost. In these cases, sufficient auxiliary fuel must be provided to completely heat the incinerator gas to the ignition temperature and this heat can be further utilized. Steam can be generated if there is a need for process steam. If steam is not needed, moderate to large size incinerators have a heat exchanger between the incoming gases and the hot combustion products to preheat make-up air. Because of the temperatures involved, heat-transfer surfaces are generally of alloy construction or ceramic materials. The latter can be damaged by thermal shock if sudden step changes in operation occur, whereas alloys may be attacked by corrosive gases such as halogens and sulfates if present. Typical heat-recuperation devices are finned gas exchangers, ceramic heat wheels, and Ljungstrom air preheaters.

Turbulence is enhanced in direct-flame incinerators by imparting shear from gas streams directed in opposing directions or having differing velocities such as jets introduced with cyclonic flow. Baffles have the same effect, whereas radiant refractory surfaces in contact with the gas speed combustion through surface catalysis. For incinerator design, the chemical composition of the waste gas must be known, and the heat of combustion and gas temperature rise must be calculated in order to choose refractories having adequate service temperature. Safety devices should be provided to shut off flows in the event of flame failure, to provide purging prior to restarting and ignition, and to limit peak temperatures.

Flares. Flares are used for burning concentrated gases that support combustion such as hydrocarbon blowdown gases, tank venting, and emergency releases. These are located well above other structures. Pilot flames must be arranged to ensure ignition of all combustibles, especially when the flare handles only emergency releases. Design details to be considered are (1) pilot flames that stay lit and can be relit even in very high winds or heavy rains, (2) flare height and location to protect both personnel and the surroundings, (3) protection against flashback to the process, (4) production of explosive mixtures in the flare pipe from intermittent flows, and (5) entrained combustible liquid removal from the flare gas to prevent burning liquid droplets falling from the flare.

For environmental reasons, burning should be smokeless. Long-chain and unsaturated hydrocarbons crack in the flame producing soot. Steam injection helps to produce clean burning by eliminating carbon through the water gas reaction. The quantity of steam required can be as high as 0.05–0.3 kg steam per kg of gas burned. A multijet flare can also be used in which the gas burns from a number of small nozzles parallel to radiant refractory rods which provide a hot surface catalytic effect to aid combustion.

When burning hazardous vapors requiring destruction of the molecular species to less than 5–10 ppmv, an open flare may not maintain the combustible gases at a sufficiently high temperature for a sufficient period of time to meet such requirements. Either pretesting of flares should be made, or the flare should be enclosed in an open-ended refractory chamber to maintain combustion temperatures. For occasional emergency releases, an enclosure built of refractory brick gives inadequate destruction until the refractory is heated to high temperatures. Maintaining a refractory lining at operating temperature with an auxiliary fuel over a long period of time can be very energy consuming.

BIBLIOGRAPHY

"Incinerators" in *ECT* 3rd ed., Vol. 13, pp. 182–206, by B. B. Crocker, Monsanto Co., and R. C. Bailie, West Virginia University.

1. P. F. Fennelly, *The Role of Incineration in Chemical Waste Management*, American Institute of Chemical Engineers, New York, 1986.
2. *Hazardous Waste Incineration—A Resource Document*, American Society of Mechanical Engineers, New York, 1988, p. ix.
3. U.S. Environmental Protection Agency, *Hazardous Waste Incineration: Questions and Answers*, EPA 530-SW-88-018, Washington, D.C., Apr. 1988, p. 1.
4. C. R. Dempsey and E. T. Oppelt, *J. Air Waste Manage. Assoc.* **43**, 25–73 (1993).
5. J. V. L. Kiser, *Waste Age*, 139 (Nov. 1990).
6. C. C. Lee, *Medical Waste Incineration Handbook*, Government Institutes, Inc., Nov. 1990, pp. 2–10.
7. E. M. Steverson, *Environ. Sci. Technol.*, **25**(11), 1810 (1991).
8. K. D. Phillips, *Atlanta J.*, 10A (June 8, 1988).
9. *U.S. Code of Federal Regulations*, Title 40, Part 261, Appendix II, Washington, D.C., July 1991.
10. Ref. 9, Appendix VIII.
11. J. De Marco and co-workers, *Incineration Guidelines*, PHS Publication 2012, Cincinnati, Ohio, 1969.
12. J. S. Lighty, G. D. Silcox, D. W. Pershing, and V. A. Cundy, *Proceedings from APCA 81st Annual Meeting and Exhibition, June 19–24*, paper 88-17.5, Dallas, Tex., 1988.
13. L. Tognotti, M. Flytzani-Stephanopoulos, A. F. Sarofim, H. Kopsinis, and M. Storekides, *Environ. Sci. Technol.* **25**, 104 (1991).
14. B. R. Keyes and G. D. Silcox, *Environ. Sci. Technol.* **28**, 840–849 (1994).
15. F. S. Larsen, G. D. Silcox, and B. R. Keyes, *Combustion Sci. Technol.* (1994) (in press).
16. H. Henein, "Radial Segregation in Rotary Kilns: A Review." presented at the *Third TIRKUA Symposium*, Ocho Rios, Jamaica, May 15, 1987.
17. J. R. Ferron and D. K. Singh, *AIChE J.* **37**, 747–758 (1991).
18. F. S. Larsen, *The Thermal Treatment of Contaminated Soils and the Incineration of Waste Fuels*, PhD dissertation, University of Utah, Salt Lake City, 1994.
19. H. C. Hottel and A. F. Sarofim, *Radiative Transfer*, McGraw-Hill Book Co., Inc., New York, 1967, pp. 453–459.
20. G. D. Silcox and D. W. Pershing, *J. Air Waste Manage. Assoc.* **40**, 337–344 (1990).
21. W. D. Owens, G. D. Silcox, J. S. Lighty, X. Deng, and D. W. Pershing, *Combust. Flame* **86**, 101 (1991).
22. S. H. Tscheng and A. P. Watkinson, *Canadian J. Chem. Eng.* **57**, 433 (1979).
23. J. P. Gorog, T. N. Adams, and J. K. Brimacombe, *Metallur. Trans. B.* **13B**, 153 (1982).
24. H. Kobayashi, J. B. Howard, and A. F. Sarofim in the *Proceedings of the Sixteenth Symposium (International) on Combustion*, The Combustion Institute, Pittsburgh, Pa., 1977, pp. 411–425.
25. D. B. Anthony, J. B. Howard, H. C. Hottel, and H. P. Meissner, *Proceedings of the Fifteenth Symposium (International) on Combustion*, the Combustion Institute, Pittsburgh, Pa., 1975, pp. 1303–1317.

26. W. D. Owens and co-workers, *J. Air Waste Manage. Assoc.* **42**, 681 (1992).

27. J. P. Reynolds, R. R. Dupont, and L. Theodore, *Hazardous Waste Incineration Calculations: Problems and Software*, John Wiley & Sons, Inc., New York, 1991.

28. J. A. Miller and C. T. Bowman, *Prog. Energy Combust. Sci.* **15**, 287 (1989).

29. W. D. Owens, *Hazardous Waste Incineration in a Rotary Kiln*, Ph.D. dissertation, University of Utah, Salt Lake City, 1991.

30. D. A. Tillman, W. R. Seeker, D. W. Pershing, and, K. DiAntonio, *Remediation*, 251 (summer 1991).

31. R. C. Flagen and J. H. Seinfeld, *Fundamentals of Air Pollution Engineering*, Prentice Hall, Inc., Englewood Cliffs, N.J., 1988, pp. 358–390.

32. W. P. Linak and J. O. L. Wendt, *Prog. Energ. Combust. Sci.* **19**, 145–185 (1993).

33. B. K. Gullett, P. M. Lemieux, and J. E. Dunn, *Environ. Sci. Technol.* **28**, 107–118 (1994).

34. R. B. Bird, W. E. Stewart, and E. N. Lightfoot, *Transport Phenomena*, John Wiley & Sons, Inc., New York, 409 (1960).

35. J. W. Moore and R. G. Pearson, *Kinetics and Mechanism*, 3rd ed., John Wiley & Sons, Inc., New York, 1981, pp. 159–181.

36. S. W. Benson, *Thermochemical Kinetics*, 2nd ed., John Wiley & Sons, Inc., New York, 1976, p. 98.

37. D. S. Duvall and W. A. Rubey, "Laboratory Evaluation of High Temperature Destruction of Kepone and Related Pesticides," EPA-600 2-76-299, NTIS, Springfield, Va., Dec. 1976.

38. K. C. Lee, J. L. Hansen, and D. C. Macauley, "Predictive Model of the Time-Temperature Requirements for Thermal Destruction of Dilute Organic Vapors," *72nd Annual APCA Meeting*, Cincinnati, Ohio, June 1979.

39. J. L. Graham, W. A. Rubey, and B. Dellinger, "Determination of Thermal Decomposition Properties of Toxic Organic Substances," *Summer National Meeting*, AIChE, Cleveland, Ohio, Aug. 1982.

40. E. E. Kahlil, D. B. Spalding, and J. H. Whitelaw, *Int. J. Heat Mass Trans.* **18**, 775–791 (1975).

41. J. M. Beer and N. A. Chigier, "Combustion Aerodynamics," Halsted Press, New York, 1972, pp. 125–126.

42. M. W. Lister, *Can. J. Chem.* **34**, 479 (1956).

43. C. N. Satterfield, *Heterogeneous Catalysis in Practice*, McGraw-Hill Book Co., Inc., New York, 1980, pp. 229–231.

44. W. M. Edwards, "Mass Transfer and Gas Absorption," in R. H. Perry, D. W. Green, and J. O. Maloney, eds., *Perry's Chemical Engineers' Handbook*, 6th ed., McGraw-Hill Book Co., Inc., New York, 1984.

45. J. F. Zemaitis, D. M. Clark, M. Rafal, and N. C. Scrivner, *Handbook of Aqueous Electrolyte Thermodynamics*, AIChE, New York, 1986.

46. M. Rafal and S. J. Sanders, "The ProChem System for Modeling Simulation of Aqueous Systems," *Proceedings of the Second International Airline House Conference on Aqueous Systems*, Warrenton, Va., May 10–14, 1987.

47. S. Gordon and B. J. McBride, "Computer Program for Calculation of Complex Chemical Equilibrium Composition, Rocket Performance, Incident and Reflected Shocks, and Chapman-Jouget Detonations," NASA SP-273, Interim Revision, NTIS, Springfield, Va., Mar. 1976.

48. R. J. Kee, J. A. Miller, and T. H. Jefferson, "CHEMKIN: A General-Purpose, Problem-Independent, Transportable, FORTRAN Chemical Kinetics Code Package," DOE, Sandia Laboratory SAND80-8003, NTIS, Springfield, Va., Mar. 1980.

R. Bertrum Diemer, Jr.
Thomas D. Ellis
E. I. du Pont de Nemours & Co., Inc.

Geoffrey D. Silcox
JoAnn S. Lighty
David W. Pershing
University of Utah

INCLUSION COMPOUNDS

The history of inclusion compounds (1,2) dates back to 1823 when Michael Faraday reported the preparation of the clathrate hydrate of chlorine. Other early observations include the preparation of graphite intercalates in 1841, the β -hydroquinone H_2S clathrate in 1849, the choleic acids in 1885, the cyclodextrin inclusion compounds in 1891, and the Hofmann's clathrate in 1897. Later milestones of the development of inclusion compounds refer to the tri-*o*-thymotide benzene inclusion compound in 1914, phenol clathrates in 1935, and urea adducts in 1940.

Perhaps the first turning point in understanding the nature of inclusion compounds occurred in the late 1940s with pioneering x-ray crystal studies on the β -hydroquinone clathrates (3). At this time the term clathrate, derived from the Latin word *clathratus* meaning enclosed by the bars of a grating, was coined to describe the principle of these systems possessing cage-like structures for the accommodation of a secondary species (4). The more general term inclusion compound (from the German *Einschlußverbindung*) was introduced in 1952 by the German chemist Cramer (5) who gave the following definition: "All these compounds namely have in common the ability to incorporate into the cavities of their own molecules, or within their lattices, other molecules of suitable size, spatially to enfold them, that is to hold them, though not by main or secondary valence forces, but mostly by physical imprisonment." A scheme for this process is given in Figure 1.

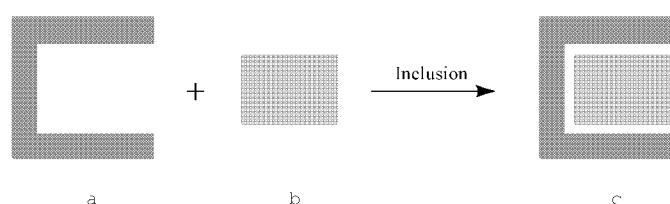


Fig. 1. The principle of formation of an inclusion compound. (a) concave host; (b) convex guest component; (c) host-guest compound.

The real breakthrough in inclusion chemistry, however, is attached to the discovery of crown compounds by Pedersen (6) in the mid-1960s. From this time, a tremendous variety of compounds and materials all having the attribute of inclusion were developed giving rise to what is called host-guest chemistry (7,8), a typical feature of which is accommodation of a complementary guest species into a concave host framework involving a molecular recognition process (9–11) such as imaged by the complementarity of a lock and a key (12) (see Fig. 1). For their pioneering studies in this field, the triumvirate Pedersen (13), Cram (14), and Lehn (15) were awarded the 1987 Nobel Prize in chemistry (16). Logically a new broad area of chemistry grew up on this basis during the last two decades. It is called supramolecular chemistry (17–20) and means the chemistry beyond the molecule (15). Here, noncovalent bonds and spatial fit between molecular individuals that form a specific supramolecular complex (inclusion complex) are in the foreground. This new direction of science and thinking where concave frameworks and containers rather than convex molecular structures are the top target and which is understood as the third important phase of chemistry, is expected to have its culminating point ahead.

Notwithstanding the immense number and great variety of inclusion compounds (21,22), all of them may be classified into three main categories (2) being either a complex, a cavitate, or a clathrate according to the criteria given in Figure 2. Typical examples for each class of inclusion compounds are the crown complexes, the calix-cavities, and the hydroquinone clathrates but in many of the recently known inclusion situations there are borderline cases treated as complex-clathrate hybrids (coordinatoclathrates or clathratocomplexes depending on the dominant inclusion character). By way of contrast, the description addition compound (adduct) may be used to the best advantage if a cavity does not exist either at the host molecule or in the lattice build-up. Inclusion compound, therefore, is the generic term of choice which refers to the presence of any not precisely defined cavity. In a more detailed topological characterization (23), there are two-dimensional open intercalates (layer- or sandwich-type inclusions), one-dimensional open channel inclusions (tubulates), and totally enclosed cage inclusions (cryptates).

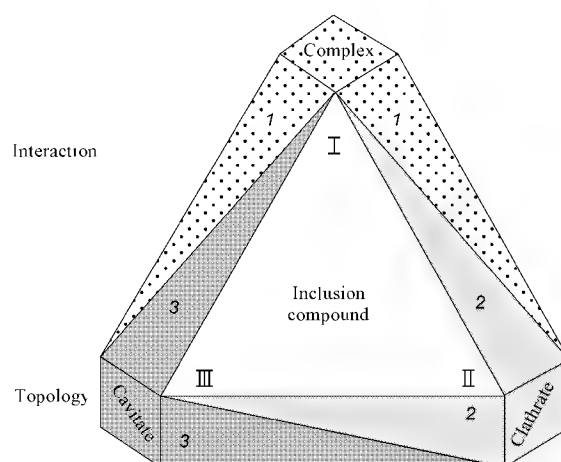


Fig. 2. Classification nomenclature of host-guest type inclusion compounds, definitions and relations: (1) coordinative interaction, (2) lattice barrier interaction, (3) monomolecular shielding interaction; (I) coordination-type inclusion compound (inclusion complex), (II) lattice-type inclusion compound (multimolecular extramolecular inclusion compound, clathrate), (III) cavitate-type inclusion compound (monomolecular intramolecular inclusion compound) (2).

Intramolecular Cavity Inclusions: Cavitates

Crown Macroring Inclusion Compounds (Coronates). Prototypical crown macro rings are cyclic oligoethers such as given by formulae (1–4) (Fig. 3). Inside the ring they make available a negatively polarized cavity capable of accommodating metal ions to form crown cation inclusion

complexes (coronates) (24–26). In particular alkali and alkaline earth metal ions that match the ring interior in size and give rise to high ion–dipole interaction are involved and make these compounds unique (eg, Li^+ 12-crown-4 [294-93-9] (1), Na^+ 15-crown-5 [33100-27-5] (2), Ba^{2+} 18-crown-6 [17455-13-9] (3), K^+ 18-crown-6 (3), Cs^+ 21-crown-7 [33089-36-0] (4). Ammonium cations and uncharged organic molecules suitable to form crown–guest hydrogen bonds are also complexed. Oxygen donor atoms have been replaced by nitrogen, sulfur, and other heteroatoms to give hetero crown ethers (27). They are efficient complexants of soft transition metal ions (Ag^+ , Hg^{2+} , etc).

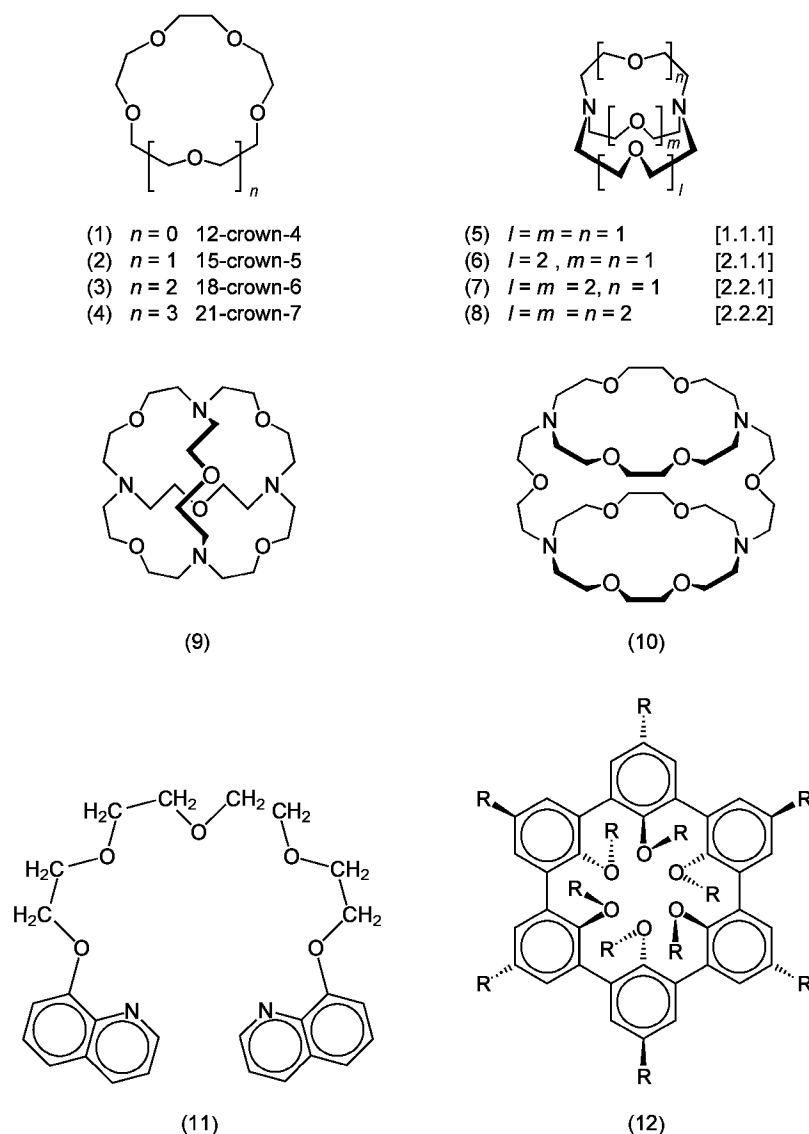


Fig. 3. Crown compounds cryptands and analogous inclusion hosts. (1–4) Crown macro rings; bicyclic cryptands (5) [37095-49-1], (6) [31250-06-3], (7) [31364-42-8], (8) [23978-09-8]; (9) spherical cryptand [56698-26-1]; (10) cylindrical cryptand [42133-16-4]; (11) a podand [57310-75-5]; and (12) a spherand [72526-85-3]. R = methyl.

Cryptates They are inclusion complexes of quasispherical analogues of crown macro rings (cryptands) having bi- or oligocyclic frameworks (28,29). Important examples of different topology are illustrated with formulae (5–8), and (9,10) (Fig. 3). Compared with the monocyclic crown hosts they provide enforced cavities, thus giving rise to increased stability and selectivity of inclusion complexes, eg, bicyclic cryptands (5–8) preferably accommodate H^+ , Li^+ , Na^+ , and K^+ , in this order. The spherical cryptand (9), (soccer ball molecule) with tetrahedral orientation of nitrogen atoms strongly complexes the ammonium cation; the cylindrical cryptand (10) is well suited to give dinuclear cryptates with size matching diammonium cations.

Podates Acyclic analogues of crown ethers coronands and cryptands (podands, eg, (11) (30) are also capable of forming inclusion compounds (podates) with cations and uncharged organic molecules, the latter being endowed with a hydrogen bond functionality. Podates normally are less stable than coronates and cryptates but have favorable kinetics.

Inclusions of Other Crown Analogues. A variety of crown analogues and hybrid modifications (24–28) with other topological features (lariat ethers (31,32), octopus molecules (33), spherands (eg, (12) (34), torands (35)) including chiral derivatives (36) have been prepared and demonstrated to show particular inclusion properties such as chiroselective inclusion (Fig. 4) (37) or formation of extremely stable complexes (K_s^+ – (Li^+) for (12) $>7 \times 10^{16}$ in CDCl_3 saturated with D_2O at 25°C) (34).

For thermodynamic (stability constants) and kinetic data involving crown-type inclusion complexes see References r38 and r39; structural results in References r40–r42 (see also CHELATING AGENTS).

Cyclophane Host Inclusion Compounds. Cyclophane-type hosts are determined by a polycyclic framework composed of rigid aromatic groups and linkers rather than by the presence of a set of donor heteroatoms typical of crown compounds and their analogues (43–45). Today, they represent the central class of synthetic receptors in molecular recognition and inclusion. Inclusion phenomena involve all kinds of *ortho*-, *meta*-, and *para*-bridged aromatic host topologies including macro rings, intercalands, pockets, open vessels, and macrocages. They may have *exo*- or *endo*-polar groups and additional functions for solubilization and/or complexation. Typical examples of this host design and of respective inclusion compounds are given by formulae (15–18) (Fig. 5) showing that hydrogen bonds, π – π -stacking interactions, and steric fit all play important roles. Complementary guest species to be included are shapely molecules of suitable size involving hydrogen bond donors and acceptors or aromatic compounds. Certainly, the case of inclusion having the highest level of imprisonment is shown with formula (17). According to the topological image, hosts of this type have been called carcerands and their inclusion compounds are named carceplexes (46). Incarceration of the guest is permanent here. Escape from the host interior is impossible without cleavage of covalent bonds of the container shell. Other namings for particular types of cyclophane hosts based on structural features are cryptophane (aromatic cryptands) (47) or collarane (collar-like hosts) (48,49). Moreover, macrocyclic heteroaromatic analogues containing the bipyridino–paraquat building block, were designed that form charge-transfer type inclusion complexes with aromatic guests having electron-donating substituents (50). The principle has been used masterfully to make formation of self-organized host–guest rotaxanes and catenanes possible (51). Chiral analogues of cyclophane hosts for chiroselective inclusion of guest molecules are also available (44).

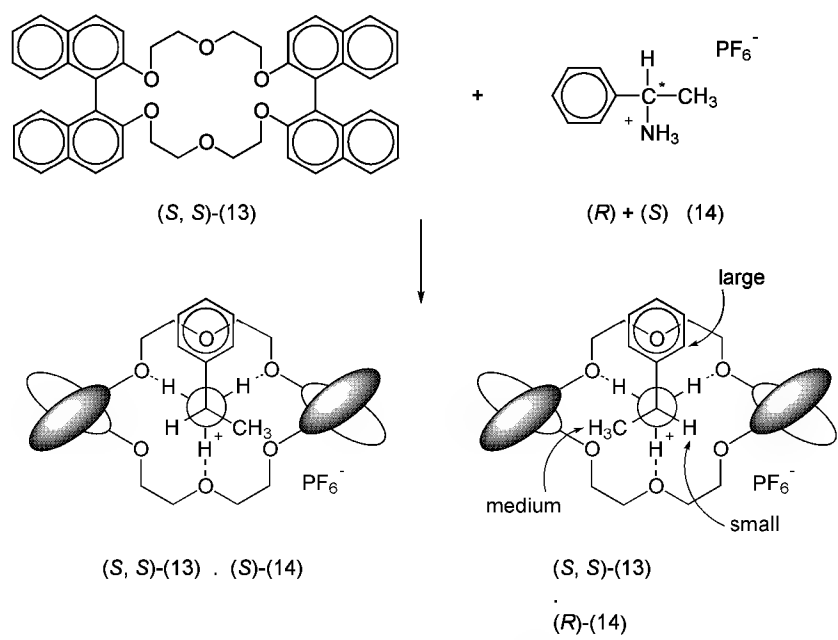


Fig. 4. Chiroselective inclusion formation of racemic 1-phenylethylammonium salt ((R/S)-**14**) using optically active crown compound ((S,S)-**13**) [53955-48-9]. The diastereomeric inclusion complex (S,S)-**13** · (R)-**14** is more stable than (S,S)-**13** · (S)-**14** (top views, dotted lines represent hydrogen bonds) thus making enantio separation of (R/S)-possible-(**14**) possible. Large, small, and medium refer to phenyl, hydrogen, and methyl, respectively.

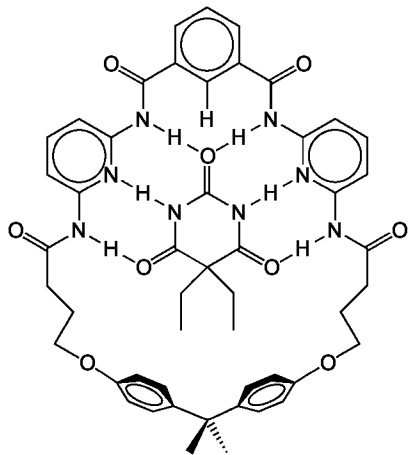
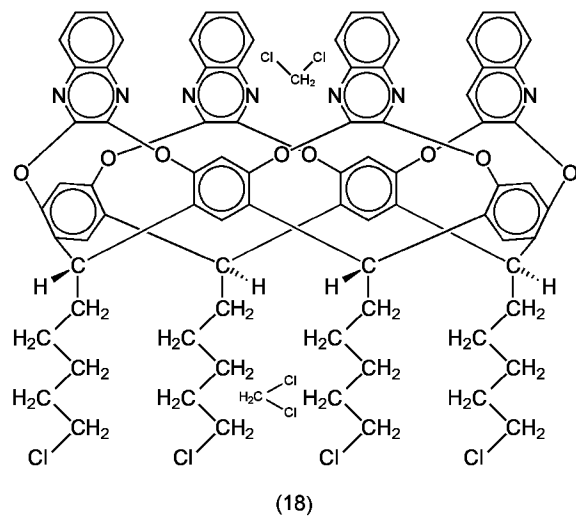
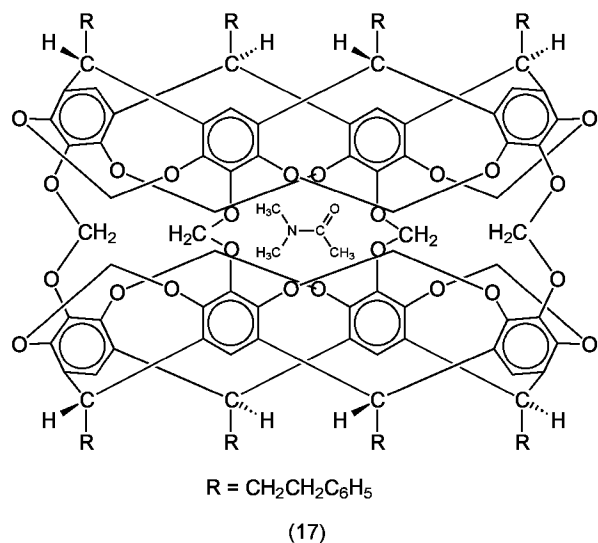
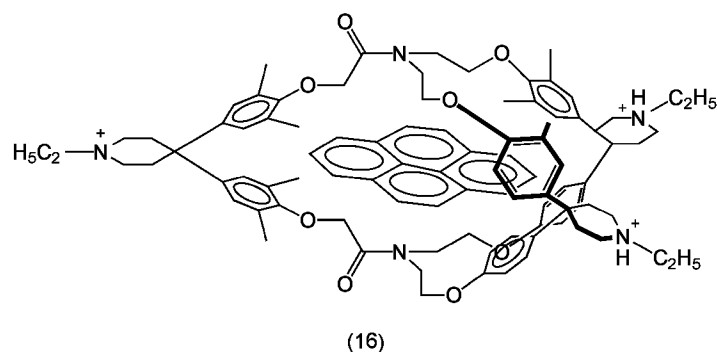
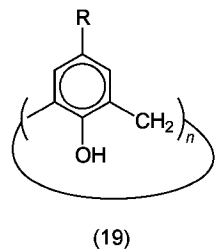
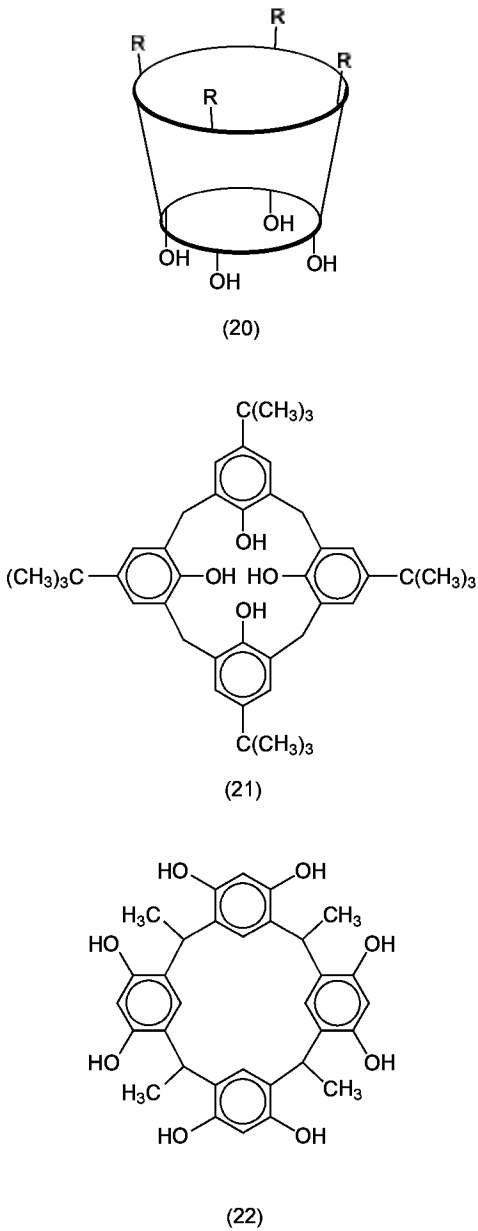


Fig. 5. Cyclophane-type inclusion compounds of different varieties. The guest component is shaded.



Calixarene Inclusion Compounds. Calixarenes are a particular class of metacyclophane hosts bearing protonizable hydroxy groups (52,53). In the original sense, they are cyclic oligomers produced by condensation of *p*-substituted phenols with formaldehyde (19). The class name calixarene was chosen due to the characteristic cone- or calix-like conformation (20), especially of such molecules with four aromatic moieties (calix[4]arene); the higher homologues (calix[6]arene and calix[8]arene) are more flexible. An enormous variety of such compounds with different ring sizes and substituents have been synthesized since the early 1970s. Although broad in variation, most of these compounds derive from the two prototypes *p*-*t*-butylcalix[4]arene [60705-62-6] (21) and *C*-methylcalix[4]resorcinarene [74708-10-4] (22) having hydroxy groups either at the lower or upper rims of the calix. Compound (21) is the one-pot cyclocondensation product of 4-*t*-butylphenol [98-54-4] with formaldehyde, (22) of resorcinol with acetaldehyde. Special calixarenes where the *t*-butyl groups of (21) (and higher ring analogues) are removed and replaced by other functional groups or more sophisticated constructions involving different phenolic or resorcinol units require a stepwise synthesis. The same is true for compounds with bridged *p*-positions and systems where two calixarenes are connected via their *p*-positions. Note that carcerate (17) and cavitate (18) of Figure 5 are based on calix[4]resorcinarene (22) as the supporting frame. Modification of the phenolic hydroxyl makes special calixarenes available (54).





Many calixarenes show a spontaneous ability to retain the solvent from which they are crystallized (eg, benzene, toluene, xylene, anisole, chloroform) to give a solid state inclusion compound (52,53). They may be divided into three main categories: intramolecular, cage-type, and intermolecular. In the 1:1 inclusion compound between (21) and toluene, the guest is accommodated into the hydrophobic pocket of the host cone conformation. Anisole yields a 2:1 cage inclusion compound with (21) where the guest is enclosed in the cage formed by two facing intramolecular cavities of two calixarenes. In the case of the 1:1 inclusion compound between the bulkily substituted *p*-(1,1,3,3-tetramethylbutyl)calix[4]arene and toluene partial filling of the cone cavity by *t*-butyl groups of neighboring hosts is observed. On the other hand, the bulky chains give rise to interstitial lattice space including the guest, a clathrate type inclusion compound. Reference r53 depicts these three inclusion compounds described.

Amalgamation of structural units typical of crowns and calixarenes has led to the development of calixpodands, calixcrowns, and calixspherands (55). Naturally they behave as cation complexants rather than inclusion hosts for uncharged molecules.

Cyclodextrin Inclusion Compounds. Cyclodextrins comprise a family of cyclic oligosaccharides obtained from starch by enzymatic degradation (56,57). Three of them, the so-called α -, β -, and γ -cyclodextrins (23) Fig. 6a), are of great importance. They are composed of 6, 7, and 8 glucose units, respectively, and form cone structures similar to the calixarenes (Fig. 6b). Thus, the cyclodextrins may be understood as natural analogues of the artificial calixarenes. Dimensions of the cone-shaped molecular cylinders of α -, β -, and γ -cyclodextrin are given in Table 1. In contrast with the calixarenes, the cavities are lined on both rims with hydroxyl groups. The interior of the cavity is hydrophobic whereas the outside is hydrophilic. These unique properties explain some of the unusual features of the cyclodextrins. They are easily soluble in water and form inclusion compounds with a wide variety of guest species in solution and in the solid state. Nobel gases, paraffins, alcohols, carboxylic acids, aromatic dyes, benzene derivatives, salts, etc, are included, just to name a few of a long list of potential substances; the only obvious requirement is that the guest must fit into the cavity even if only partly. It has been confirmed via x-ray crystallography that the central cavities of α - and β -cyclodextrin contain two and nine water molecules, respectively. On formation of inclusion compounds, the water molecules are displaced by the guests.

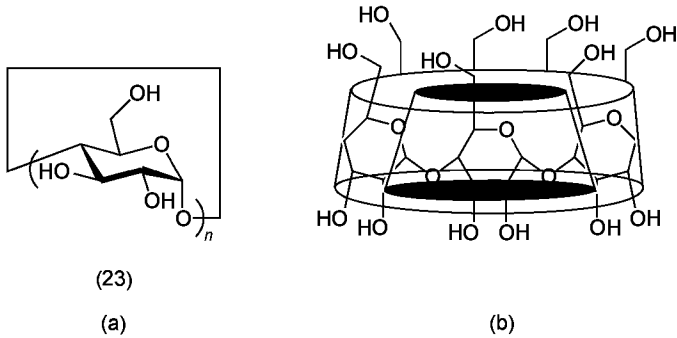


Fig. 6. Cyclodextrins: (a) formula representation for α -cyclodextrin [10016-20-3] (23)a, $n = 6$, β -cyclodextrin [7585-39-9] (23)b, $n = 7$, and γ -cyclodextrin [17465-86-0] (23)c $n = 8$. (b) Three-dimensional cone structure of β -cyclodextrin (23)b.

Table 1. Dimensions of Cone-Shaped Molecular Cylinders of Cyclodextrins

Cyclodextrin	Molecular weight	Solubility in water, g /100 mL	$[\alpha]_D^{25}$	Diameter of cavity, ^a nm	Diameter of outer periphery, ^a nm
α-cyclodextrin	972	14.5	150.5 ± 0.5	0.47–0.52	1.46 ± 0.04
β-cyclodextrin	1135	1.85	162.5 ± 0.5	0.60–0.64	1.54 ± 0.04
γ-cyclodextrin	1297	23.2	177.4 ± 0.5	0.75–0.83	1.75 ± 0.04

^a Measured with CPK-models. The height of all cyclodextrins is 0.79–0.80 nm.

Packing of the cyclodextrin molecules (α, β, γ) within the crystal lattice of inclusion compounds (58,59) occurs in one of two modes, described as cage and channel structures (Fig. 7). In channel-type inclusions, cyclodextrin molecules are stacked on top of one another like coins in a roll producing endless channels in which guest molecules are embedded (Fig. 7a). In crystal structures of the cage type, the cavity of one cyclodextrin molecule is blocked off on both sides by neighboring cyclodextrin molecules packed crosswise in herringbone fashion (Fig. 7b), or in a motif reminiscent of bricks in a wall (Fig. 7c).

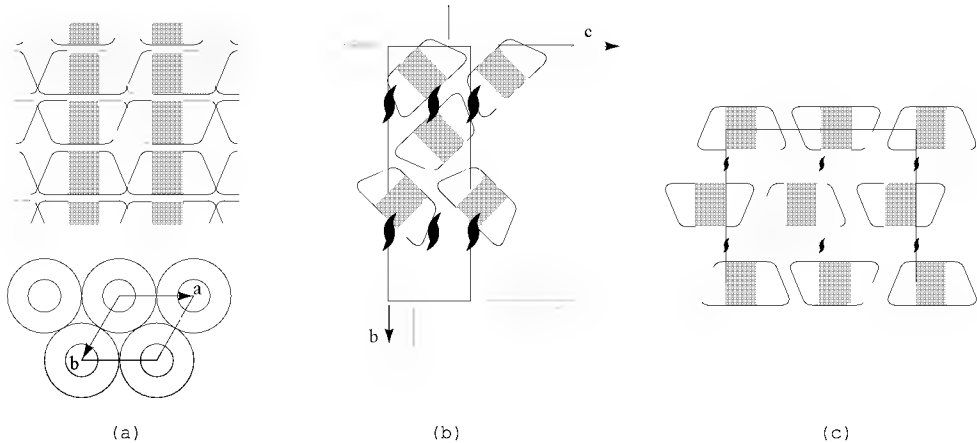


Fig. 7. Schemes of crystalline cyclodextrin inclusion compounds: (a) channel type; (b) cage herringbone type; (c) cage brick type (58).

Thermodynamics and kinetics of cyclodextrin inclusion compounds have been reported and mechanisms have been suggested by which the guests are enclosed in the cyclodextrin ring (60). These factors and others have led to structural modifications involving the addition of flexible and rigid caps to one end of the cyclodextrin cylinder (61) or linkage of two cyclodextrins (duplex cyclodextrin) (62) which proved efficient in enhancement of the guest binding. More sophisticated functional systems based on cyclodextrins as models for several enzymes (63) have also been constructed.

Amylose Inclusion Compounds. Like cyclodextrins (Fig. 6a) amylose also consists of glucose units (5). However, amylose is not macrocyclic in structure and its mol wt is much higher (300,000–1 million, depending on the starch source from which it is derived). Amylose forms inclusion compounds mostly with long-chain fatty acids but also with iodine to give the well-known blue starch–iodine complex (64). These inclusion compounds of amylose apparently are of channel character (Fig. 8). In the latter case, iodine atoms are joined together to form a long, straight polyiodide chain spirally wrapped by the amylose polymer. Neither the helical configuration of the starch molecule nor the chain of iodine atoms is stable except for inclusion structures of this general type (see CARBOHYDRATES; STARCH).

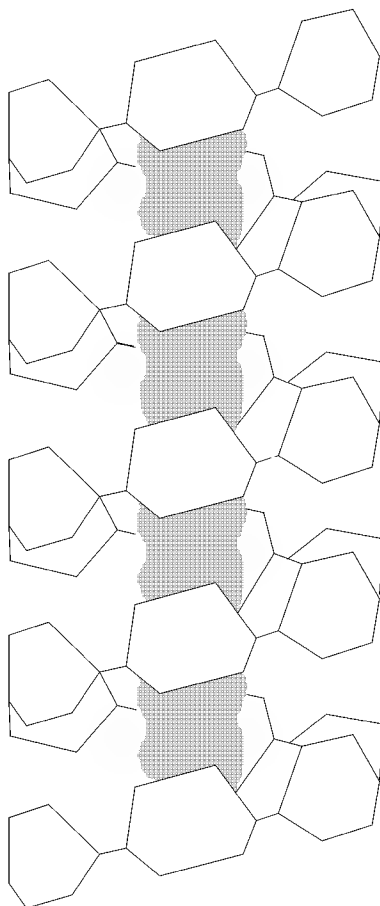


Fig. 8. Amylose inclusion compounds: scheme of the amylose helix showing the inclusion channel. The guest compound is shaded (58).

Cucurbituril Inclusion Compounds. Cucurbituril [80262-44-8] (**24**) is a nonadecacyclic cage molecule readily produced by self-assembly from urea, glyoxal, and formaldehyde (Fig. 9a) (65). The designated trivial name derives from the general resemblance of models of the molecule to a pumpkin (of botanical family Cucurbitaceae) (66). The most notable feature as regards molecular inclusion is the presence of an internal cavity of approximately 0.55 nm diameter within the rigid macrocyclic structure, to which access is provided by a 0.4 nm diameter occlusus situated among the carbonyl groups on both the top and the bottom of the molecule. The cavity inside of cucurbituril can hold small organic molecules with a particular preference for alkylammonium ions (Fig. 9b). This has been established crystallographically and in solution by nmr spectroscopy. Solid complexes of (**24**) with dye-stuffs (67) as well as the synthesis of the next lower analogue (decamethylcucurbit[5]uril) (68) have been reported.

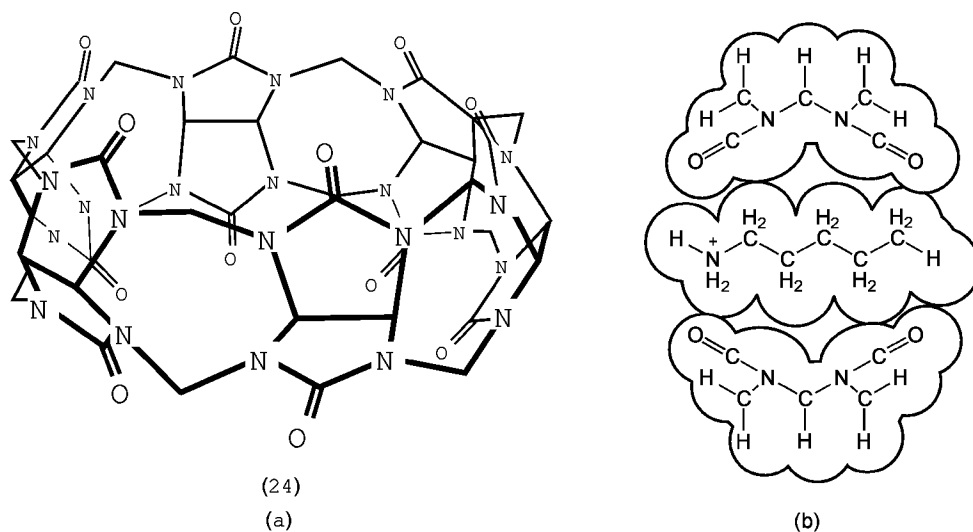
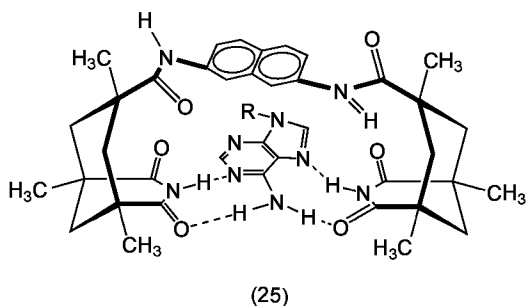


Fig. 9. Cucurbituril inclusion chemistry: (a) tridimensional structure of cucurbit[6]uril; (b) conjectured cross-sectional representation of a host-guest inclusion compound between curbit[6]uril and *n*-pentylammonium ion.

Molecular Cleft Inclusion Compounds. Nonmacrocylic hosts for the inclusion of uncharged guests that contain preorganized molecular clefts aligned with binding sites are a promising new development (69–71). Owing to structural mimicry and depending on the origin they are called molecular clefts, clips, or tweezers. One example (**25**) shows the host design and illustrates the binding and inclusion principles for adenines as guest molecules. Specific host-guest hydrogen bonds and π - π -stacking interactions are decisive factors for inclusion here. This general receptor design for the docking of guests appears to be a clever approach to making enzyme mimics and catalytic systems (11).



Anionic Guest Inclusion Compounds. Artificial inclusion hosts for anions (72–75) should contain cationic or electron-deficient binding sites to complement and neutralize the negative charge of these guests, opposite to the way cations are complexed (crown compounds). Cationic centers can be included in the covalent framework of the host (eg, ammonium, guanidinium, or phosphonium groups) and may also incorporate hydrogen bond donor groups, or (similar to metallo enzymes) cationic substructures of so-called cascade complexes may serve as anchor groups for anionic guests. Electron-deficient binding sites are provided by Lewis acid centers (boron, mercury, and tin) incorporated in the host frame. Representative examples for each category are illustrated in Figure 10. The hexaprotonated bis-tren cryptand of (26) selectively includes the linear azide anion which ideally matches the cavity and is held by the two arrays of three $N^+ - H \cdots N$ hydrogen bonds. In a cascade type of inclusion (27) two dien dicobalt substructures of a macrocyclic complex to oxalate. In (28) a fluoride anion is cryptated by a tin-containing host. Since anions display important roles in biology, inclusion chemistry of anions is promising, not least for the potential applications in medicine.

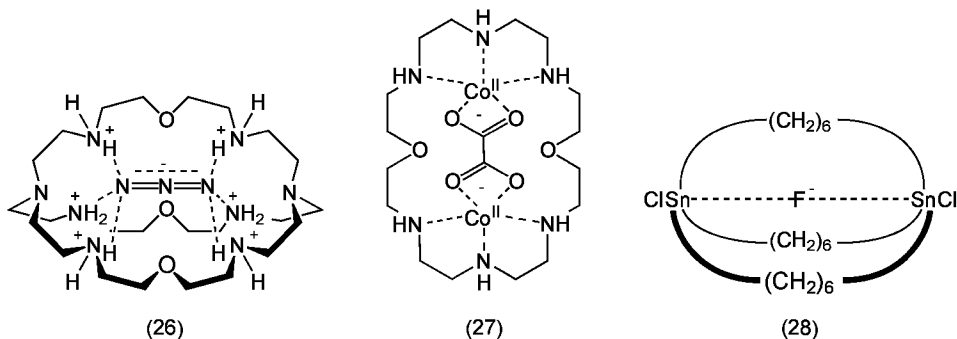


Fig. 10. Anionic guest inclusion compounds.

Extramolecular Cavity Inclusions: Lattice-Type Inclusion Compounds, Clathrates

Hydrate Inclusion Compounds. These compounds are characterized by a host lattice constructed wholly or principally from hydrogen-bonded water molecules closely analogous to that of ice (76–79). No less than four different categories can be classified although a species termed as clathrate hydrate is most frequent. Here the guests are trapped inside isolated cavities of the water matrix. The substances with which water forms clathrate hydrates are largely gases or liquids of low boiling point. This is why the clathrate hydrates are historically known as gas hydrates. Such hydrates fall into two groups (80). The first usually contains six guest molecules combined with 46 water molecules; the second contains one guest molecule for each 17 water molecules. The guests in the first group are small molecules such as Cl_2 , Br_2 , SO_2 , H_2S , CH_4 , C_2H_6 , CH_3Cl , or CH_2Cl_2 ; in the second group of gas hydrates slightly larger guest molecules are found, among them $CHCl_3$, C_2H_5Cl , CH_3I , CF_3Br , and C_3H_8 .

The structural feature common to all but a few clathrate hydrates is a pentagonal dodecahedron of oxygen atoms arising from hydrogen-bonded water molecules (Fig. 11a) (76). They are the main building blocks of the clathrate structure. Since space cannot be filled completely by any packing arrangement of dodecahedrons, some interstitial space must remain. It is in these spaces where the guest molecules are held. Actually other polyhedra (14-hedra and 16-hedra) do occur which, together with the regular 12-hedra, allow a periodical 3D arrangement. In the first group of clathrate hydrates (see above), the host framework is composed of 12- and 14-hedra in the ratio of 1 : 3 (Fig. 11b). When only the larger voids are occupied, the composition is as given ($46 H_2O \cdot 6$ guest). The second group of clathrate hydrates formed with larger guests are composed of 12- and 16-hedra. The unit cell consists of 136 water molecules oriented as dodecahedra enclosing 24 interstitial spaces, 8 of which are larger than the other 16. If only the larger voids are occupied, the composition is $136 H_2O \cdot 8$ guest. However, in both cases and if appropriately sized guests are present, both types of holes may be occupied to give mixed clathrate hydrates. Mention should also be made of the existence of alkylonium salt hydrates and amine hydrates involving other structures and more complex polyhedra than those specified above (76). Proteins also form highly hydrated crystals that may be seen as modified clathrate structures (81).

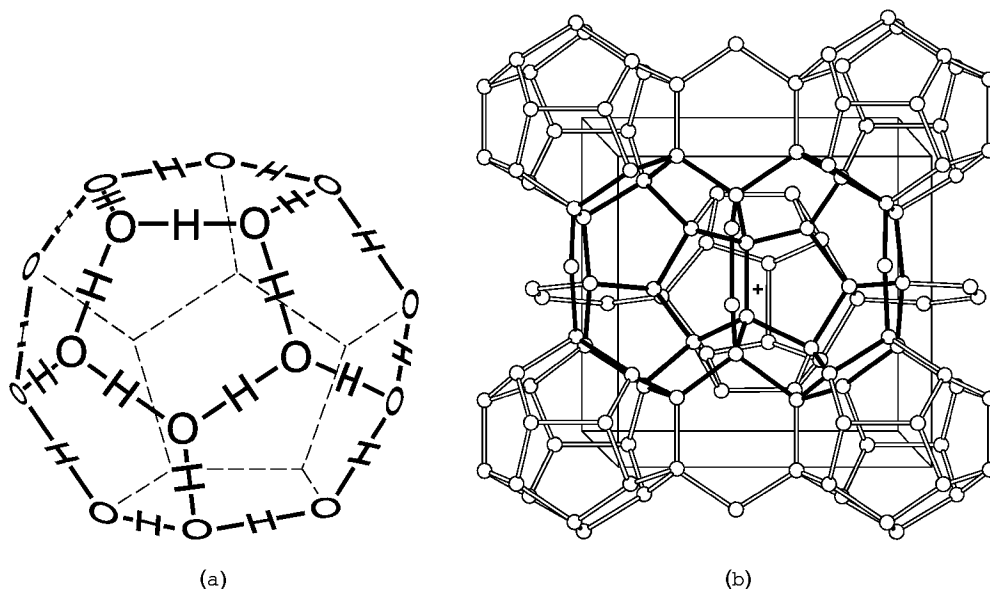


Fig. 11. Clathrate hydrates: (a) basic structural component ($H_{40}O_{20}$ pentagonal dodecahedron); (b) type I host structure (two face-sharing 14-hedra are shown with solid lines) (76).

Hofmann- and Werner-Type Inclusion Compounds. There is a wide range of clathrates having as the host component inorganic coordination compounds represented by the general formulae $M(NH_3)_2$, $M'(CN)_4$ and $M''X_2Y_4$. The first formula is typical of Hofmann-type clathrates (79,82) where M stands for Mn, Fe, Co, Ni, Cu, Zn, or Cd and M' denotes Ni, Pd, or Pt. The second specifies the Werner-type (83,84) clathrates for M'' being a divalent cation (Fe, Co, Ni, Cu, Zn, Cd, Mn, Hg, Cr), X an anionic ligand (NCS, NCO, CN, NO₃, NO₂, Cl, Br, I) and Y an electrically neutral ligand such as pyridine, substituted pyridines, or isoquinoline.

In Hofmann clathrates two guests usually belong to one host complex; the guest is benzene, thiophene, furan, pyrrole, aniline, or phenol. The prototype is $Ni(NH_3)_2Ni(CN)_4 \cdot 2C_6H_6$ (Hofmann's compound) (85). The structure consists of planar layers containing the metal atoms and the cyanide groups with the NH_3 groups protruding above and below these layers. The ammonia groups then define the void wherein the guest molecule resides. As a consequence of the essentially fixed host lattice, a selectivity towards the length of the guest molecule is provided; thus benzene can be accommodated but the longer toluene molecule cannot. Extensive studies of these host lattices have been made with respect to structure, stability, and selectivity, and several interesting modifications have been performed by using diamines as bridging units or tetrahedral $M'(CN)_4$ moieties replacing the planar complexes (86).

A wide variety of guest molecules may be trapped by the Werner-type crystalline host lattice, ranging, eg, from noble gases to condensed aromatic hydrocarbons. These clathrates may be formed from solution or by sorption. Kinetics of sorption-desorption have been studied (83).

Inclusion Compounds of Urea, Thiourea, and Selenourea. They are characteristic of channel host structures (87,88). Each channel in crystalline urea is formed by three interpenetrating spirals of molecules held together by hydrogen bonding between the nitrogen and oxygen atoms, thus forming long tubular cavities in which the guest molecules reside (Fig. 12a) (79,89,90). The spirals are randomly right- or left-handed. The inside diameter of the channels is about 0.5 nm which is just the right size to accommodate the straight-chain hydrocarbons (*n*-alkanes and *n*-alkenes). Any such substance can pack into the channels provided that the hydrocarbon has six or more carbon atoms. Urea [57-13-6] forms inclusion compounds not only with hydrocarbons but also with ethers, aldehydes, ketones, esters, carboxylic acids, alcohols, amines, nitriles, thiols, and sulfides, that exceed the lower limit of carbon atoms. Aromatic compounds also form inclusions if the benzene ring carries a long chain substituent (eg, octadecylbenzene).

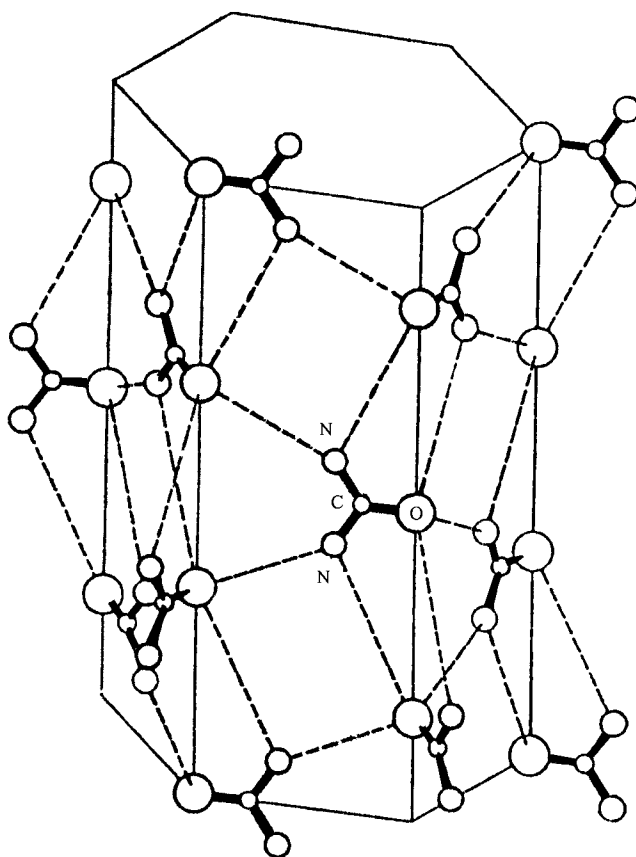


Fig. 12. Hydrogen-bonded network of an urea inclusion channel; (---) hydrogen bonds (90).

Channels in crystals of thiourea [62-56-6] (87) are comparable but, as a consequence of the larger size of the sulfur atom, have larger cross-sectional areas (0.7 nm) and can trap branched-chain, alicyclic, and other molecules of similar dimensions including polychlorinated hydrocarbons. But they do not include the straight-chain hydrocarbons that work so well with urea.

Selenourea [630-10-4] like urea and thiourea can form channel inclusion compounds (87) with a variety of hydrocarbons. Though the difference in channel diameter between thiourea and selenourea is small, selenourea seems to be much more selective for the inclusion of certain guest molecules (eg, *cis* trans isomers).

Inclusion Compounds of Phenolic Hosts. Inclusion compounds involving phenols, hydroquinone, Dianin's compound, and related molecules (91) mostly consist of cage-like structures with the guest molecules lying trapped in the cavities, a feature that initially prompted the term clathrate. The principle in forming these cages is the linkage of the OH groups of six host molecules, eg, phenol, by hydrogen bonds such that the oxygen atoms form a hexagon and alternate aromatic groups point above and below this hexagon (Fig. 13a) (79,88,92). In this way, a large open structure is created, which may be visualized as consisting of two interpenetrating cups, each cup formed by three phenolic molecules. The networks, however, do not fill the available space but remain a roughly spherical cavity for accommodation of the guests. In case of hydroquinone, both ends of the molecule are similarly involved, and a three-dimensional hydrogen-bonded network results (Fig. 13b) (4). An important feature of Dianin's compound [472-41-3] (29) in addition to the hexameric unit, is the existence of a waist halfway up the cavity formed by six inward pointing methyls whereas the lateral packing leaves no significant space (93). This has given rise to structural modifications of Dianin's compound (thiachroman, noranalogue, etc) for controlling cavity shape (91).

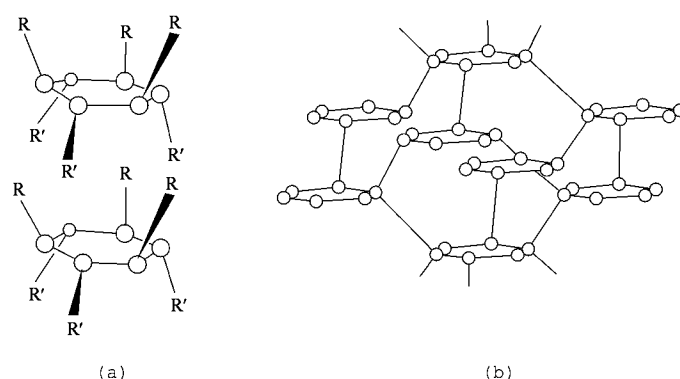
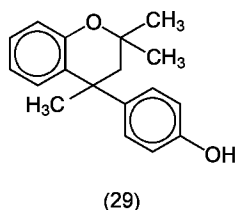


Fig. 13. Phenolic host inclusion chemistry: (a) schematic representation of the cage structure (open circles denote oxygen of OH, R corresponds to aryl part of host molecules); (b) interpenetrating cagework found in hydroquinone inclusions (91).

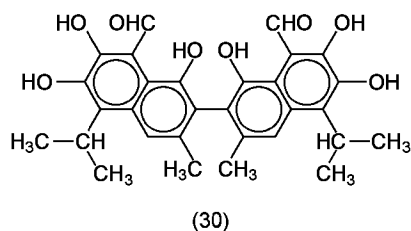


Host lattices of phenol and simple substituted phenols (91,93) usually provide an extra (relatively large) lattice space, in addition to the hydrogen-bonded cage (~ 1.5 nm effective length and 0.4–0.45 nm free diameter, and 0.45 nm free diameter, respectively, in the case of phenol). Both cavities are capable of including suitably sized guest molecules involving noble gases or other volatile species such as H_2S , SO_2 , CO_2 , CS_2 , HCl , HBr , CH_2Cl_2 , and CH_2CHF .

Examples of the hydroquinone inclusion compounds (91,93) are those formed with HCl , H_2S , SO_2 , CH_3OH , HCOOH , CH_3CN (but not with $\text{C}_2\text{H}_5\text{OH}$, CH_3COOH or any other nitrile), benzene, thiophene, CH_4 , noble gases, and other substances that can fit and remain inside the 0.4 nm cavities of the host crystals. That is, clathration of hydroquinone is essentially physical in nature, not chemical. A less than stoichiometric ratio of the guest may result, indicating that not all void spaces are occupied during formation of the framework. Hydroquinone clathrates are very stable at atmospheric pressure and room temperature. Thermodynamic studies suggest them to be entropic in nature (88).

Crystals of Dianin's compound (91,93) have been shown to accommodate more than 50 diverse guest species into the hour-glass shaped cages of about 1.1 nm length and 0.43 nm width, for example argon, I_2 , SO_2 , SF_6 , NH_3 , CH_3OH , *n*-heptanol, glycerol, di-*t*-butylnitroxide, CHCl_3 , CCl_4 , decalin, aromatic hydrocarbons, and derivatives. The host:guest ratio ranges from a low of 2:1 up to 9:1, depending on the size relationship, with 6:1 being the most frequent ratio observed. Analogues of Dianin's compound (91,93) provide increased maximum diameters of the cage (0.71 nm for analogue with a methyl group removed) and reduced cavity length (0.8 nm for 8-methyl-substituted derivative of the thia-analogue) thus giving rise to selective clathration properties.

Inclusion Compounds of Gossypol. The oligophenolic compound gossypol[303-45-7] (30), a natural product and biologically active compound, has proved to be an efficient inclusion host for a variety of low molecular weight organic substances (94). It forms clathrates of all possible types: cage, channel, and layer inclusions (95). Twenty groups of gossypol isostructural clathrates have been established. For the crystal forms of gossypol, spatial isolation of hydrophilic and hydrophobic areas is observed leading to the availability of hydrophobic and hydrophilic inclusion cavities. Gossypol itself possesses unusual polymorphism (95), consisting in the formation of eight crystallographically identified polymorphs.



Inclusion Compounds of Deoxycholic Acid (Cholic Acids). Deoxycholic acid[83-44-3] (DCA) (31) forms stable inclusion compounds with a wide variety of guests involving aliphatic, alicyclic, and aromatic hydrocarbons, alcohols, ketones, fatty acids, esters, ethers, phenols, azo dyes, nitriles, peroxides, and amines (96). The guests are generally trapped in channels that run through the host lattice (88,90). These inclusion compounds are called cholic acids. They can be grouped into three crystal systems: orthorhombic (most commonly observed and showing superior inclusion property), tetragonal, and hexagonal (96). The characteristic structural unit of the orthorhombic and tetragonal inclusion crystals is a bilayer developed by two rows of head-to-tail hydrogen-bonded molecules that are interconnected via hydrogen bonds. They assemble, related to a 2_1 or 4_1 axis, to give rise to channel space (Fig. 14a), which may have a variable size and shape depending on the mutual positions of the bilayers. This accounts for the ability of the (orthorhombic) host lattice to accommodate guest molecules of very different dimensions. The hexagonal host lattice is characterized by the packing of hollow helices of deoxycholic acid molecules, generated by a 6_3 axis (Fig. 14b). The cavity of the helix has a diameter of about 0.4 nm, which allows the inclusion only of small size or thread-like guest molecules. According to the nature of the inner channel surface, being either hydrophobic or hydrophilic, the hexagonal host can be utilized to receive polar guest molecules while the orthorhombic one prefers apolar guests. This is in line with the packing diagrams for the two crystal systems shown in Figure 14, the guest compounds being phenanthrene and dimethyl sulfoxide–water, respectively.

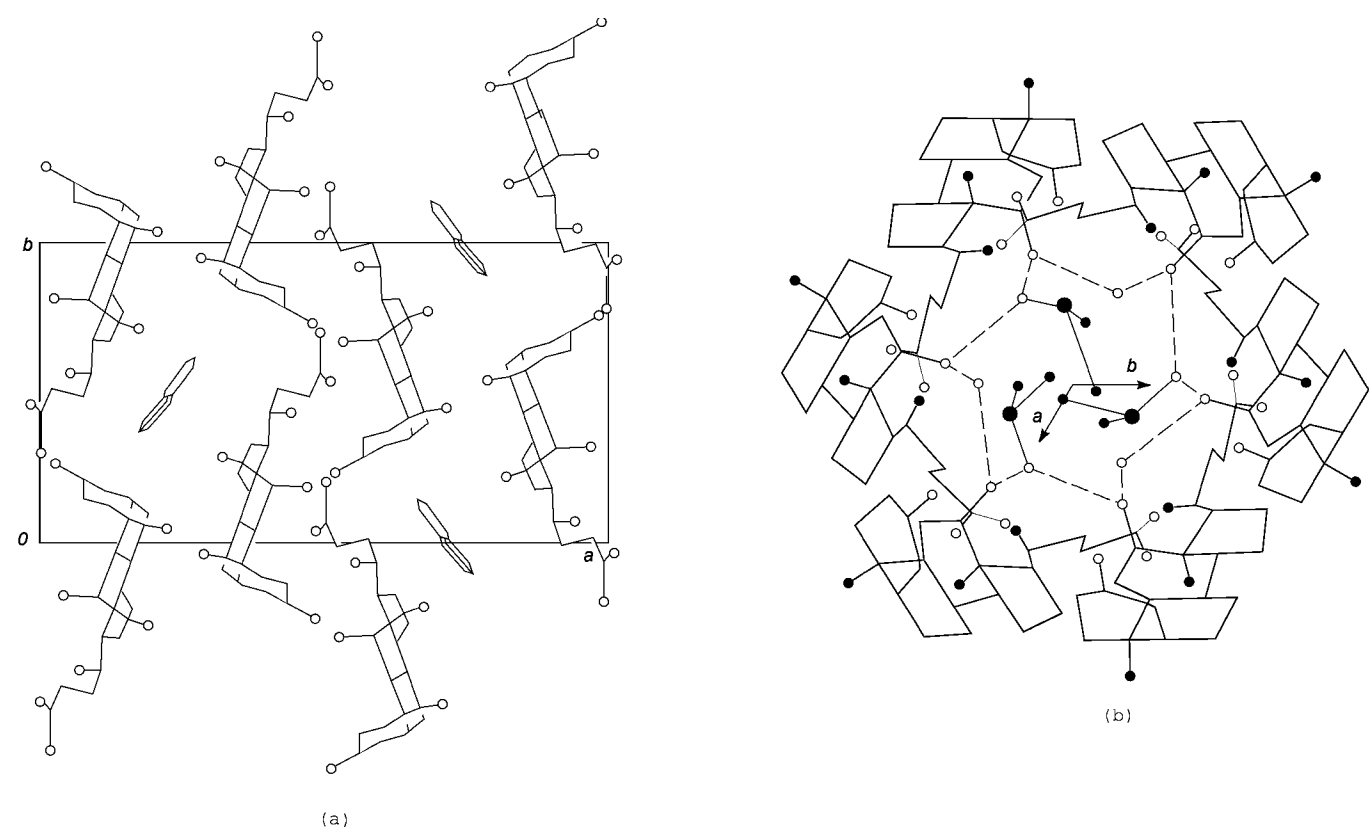
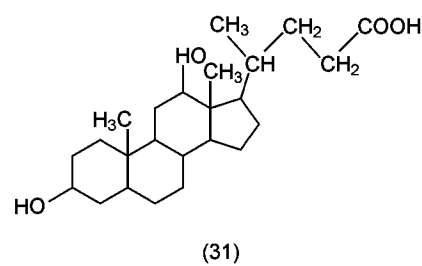


Fig. 14. Choleic acid inclusion chemistry: (a) crystal structure of DCA inclusion compound with phenanthrene; (b) view along a DCA inclusion helix accommodating DMSO and water guest molecules (oxygen and sulfur atoms and methyl groups are represented by open circles and large and small black circles, respectively). Hydrogen bonds are indicated by broken lines (96).



Inclusion Compounds of Macrocyclic and Oligocyclic Lattice Hosts. As a common feature, all these hosts (Fig. 15) belong to the trigonal class of symmetry and most inclusions are of channel structure.

Perhydrotriphenylene/[15074-91-6/ (32) comprises four cyclohexane chairs anellated in such a way as to give *D*₃ symmetry which makes the molecule chiral. Inclusion formation (79,97) is broad involving a great number of linear aliphatic hydrocarbons, ethers, alcohols, acids, and esters, but also branched compounds such as 2,2,4-trimethylpentane as well as cyclic compounds such as benzene, toluene, cyclohexane, dioxane, or spherical and quasispherical molecules such as CCl₄, CHCl₃, etc. It has been stated that it took longer to find a simple compound which would not form an inclusion than forming one. The stoichiometry of perhydrotriphenylene inclusions is not, in general, simple or even-numbered, typical of a channel structure with different repeat distances for the occluded guests. In case of the *n*-heptane inclusion compound c(98), the structure is composed of stacks of superimposed host molecules leaving channel cavities of about 0.5 nm diameter that contain the guests.

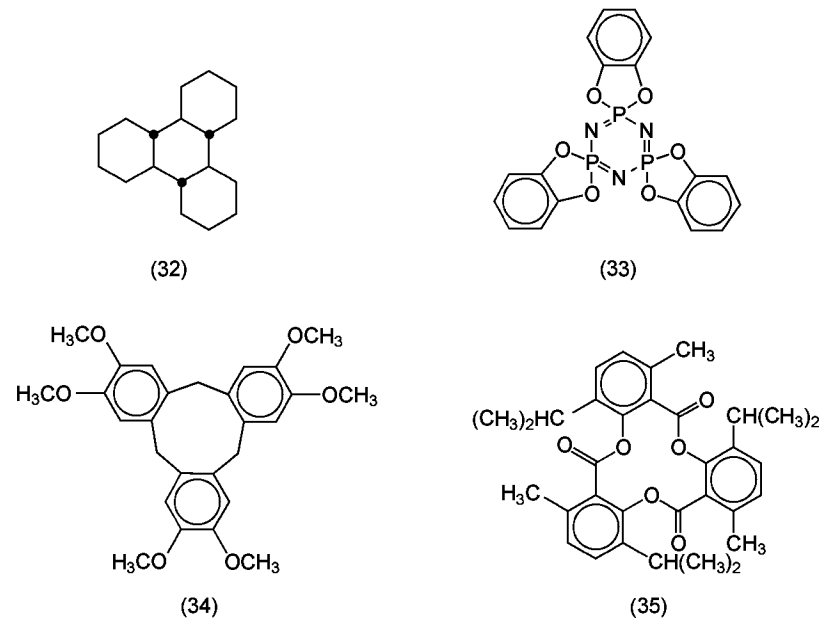


Fig. 15. Macrocyclic and oligocyclic lattice hosts: perhydrotriphenylene (32); a cyclotriphosphazene (33); cyclotrimeratrylene (34); tri-*o*-thymotide (35).

Hosts of cyclotriphosphazene type are tribladed paddle wheel-shaped molecules (see Fig. 16) of which [311-03-5] (33) is the prototype and other substitution products between hexachlorophosphazene and aromatic diols or diamines are structural variations. The cyclotriphosphazenes, eg, (33), have been shown to form inclusion compounds with a wide variety of guest molecules (99,100). Examples include aliphatic and aromatic hydrocarbons, ethers, esters, ketones, nitriles, carbon disulfide, and ethanol. The guest molecular dimensions vary from those of small molecules (see above) to larger species such as decalin, norbornadiene, or short polymer molecules. A particular feature of (33) is that inclusions are formed both by crystallization from the guest solvent and by direct exposure of the host to the vapor or liquid guest. Replacement of the included guest has also been demonstrated. A typical host lattice of (33) consists of alternating layers of molecules. Superposition of the layers leads to tunnels (0.45–0.5 nm in diameter) in which the guests reside (Fig. 16). However, there are also structures with guests trapped in cavities rather than in tunnels.

Cyclotrimeratrylene[1180-60-5] (34) is a cyclocondensation product of veratrole with formaldehyde. It possesses a stable trigonal crown conformation and forms crystalline inclusion compounds with benzene, chlorobenzene, toluene, thiophene, decalin, chloroform, acetone, carbon disulfide, acetic acid, and butyric acid (101). Their structures consist of columns of cyclotrimeratrylene molecules that are not amenable to close packing and provide channels into which the guests are accommodated. A number of modified host structures, derived from prototype (34) have been prepared. The hexaphenol analogue cyclotricatechylene also yields well-defined channel inclusions (101). They involve mostly polar guests and the structures are held by hydrogen bonding.

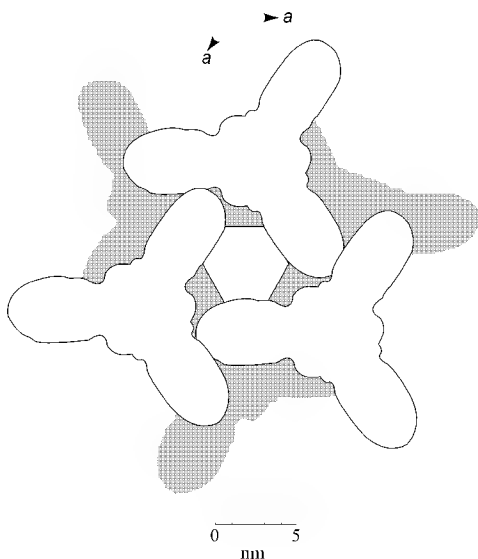


Fig. 16. van der Waals boundaries of (33) in hexagonal inclusion compounds (100).

Tri-*o*-thymotide[2281-45-0] (35) and its derivatives are cyclocondensation products of salicylic acids, and adopt an asymmetric propeller-like conformation. Inclusion compounds of thymotides, involving benzene, aliphatic ethers, alkyl halides, alcohols, and ketones, crystallize in a variety of space groups and display different modes of guest accommodation: cages, quasiuniform channels, and channels of noticeably variable section (79,102). The parameters governing the crystal form include primarily the size of the guest and, to a lesser extent, the chemical nature and the experimental conditions. It has been observed that guest molecules of length less than 0.9 nm give rise to cage-type crystals, whereas those of greater length are accommodated in uniform section channels; larger molecules (eg, stilbene) induce crystallization in channels of variable section. A host:guest stoichiometric ratio of 2:1 is usual but not general. Once formed, the inclusion compounds of (35) prove to be thermally stable. Tri-*o*-thymotide clathrates are reported to undergo spontaneous resolution, thus being attractive optical resolving agents for included guest species (102). Trianthranilides, the nitrogen analogues of (35), have also been synthesized (103). They exist in propeller and helical conformations. Their inclusion properties are less studied.

Designed Organic Host Lattices. As contrasted with the previous clathrate hosts, which have mostly been observed accidentally, these new hosts are designed compounds unrelated to any known host lattice but which would be expected to act as host lattices (2). A collection of these compounds is shown in Figure 17. They follow different design strategies to oppose the usual close packing in crystals and to create rigid open host structures.

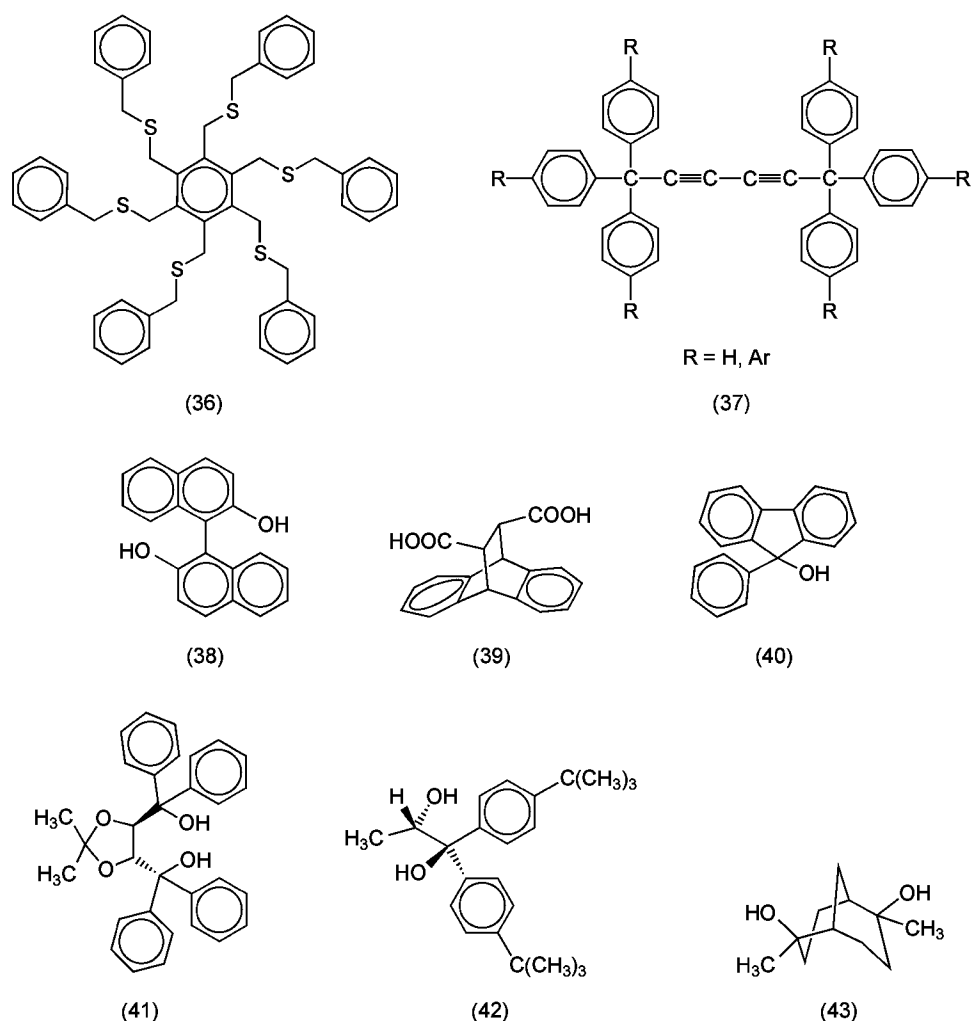


Fig. 17. Prototypical host molecules based on more recent clathrate strategies.

Thus, by analogy with the hydrogen-bonded hexamer unit present in most of the phenol-type clathrates (see Fig. 13), the synthesis of molecules with sixfold symmetry was undertaken, and a new range of inclusion hosts, named hexahosts was discovered (93,104). They feature benzenoid compounds with six substituents such as SAr , CH_2OAr , CH_2SAr , and $\text{CH}_2\text{SCH}_2\text{Ar}$, where Ar is an aromatic nucleus. The prototype of this vast family of molecules endowed with inclusion properties is hexabis(benzylthiomethyl)benzene [61040-51-5] (36). Guest species of hexahost clathrates mainly involve aromatic and alicyclic hydrocarbons, aryl and alkyl halides, dioxane, and acetone. A special merit of this host-type is that the inclusion cavity is easy to tune to the geometric and steric requirements of the guest enclosure by altering the bulk of the side arms. Using a central naphthalene moiety instead of the benzenoid ring leads to corresponding octahosts (104).

Another strategy which orients along the overall molecular shape of a host is picturesquely called the wheel-and-axle host concept (105). Such hosts, eg, 1,1,1,6,6,6-hexabis(phenyl)-2,4-hexadiyne >[90507-76-9] (37) (R = H) contain a long molecular axis made of sp carbons with sp^3 carbons at each end that bear large, relatively rigid groups (106). These act as spacers which prevent the host molecules from a close-packed structure in the crystal and give rise to inclusion properties, mostly towards nonpolar aromatic and alicyclic hydrocarbon guests. Bulky frameworks that resemble a pair of scissors (107) or a roof (108) are other basic structures of the geometrical host concept (109).

Attachment of polar groups at selected positions of the bulky host frameworks led to a broad new type of clathrate hosts that follow the principle of coordinatoclathrate formation (110) which deals with concerted action of van der Waals nonpolar steric shielding and polar Coulomb attraction or hydrogen bonding by the host, thus yielding increased chemical guest selectivity and higher stability of inclusion compounds than under van der Waals conditions alone (111,112). Representative coordinatoclathrate hosts are shown, eg, 1,1'-binaphthalene-2,2'-diol [602-09-5] (38) and the dicarboxylic acid [116841-20-4] (39), and a typical packing diagram of a respective inclusion compound indicating defined host-guest interaction is illustrated in Figure 18a. Inclusion strategies via coordinatoclathration seem to be universal and prove very efficient even when relatively simple host structures, such as 9-phenyl-9-fluorenol [25603-67-2] (40), are involved (113). Guests matching the voids of the host lattice and that have groups complementary to the host functions are usually preferred and form stable inclusions (114,115). This means, eg, that size conformable alcohols and formamides or sulfoxides are favored by carboxylic hosts (112) whereas amines show a tendency for binding to hydroxy hosts (116).

Although a number of chiral hosts (eg, (38)), based on the given principles have been found good for enantioselective guest inclusion (105,106,117), the development of optically pure clathrate hosts for enantioseparations of guests is more advanced. The new strategy consists in addition of bulky and rigid substituents (clathratogenic groups) to a natural chiral compound thus giving an optically pure lattice host (118). Tartaric and lactic acids transformed in this way (41) [93379-48-7], (42) (119,120) show efficient clathration and excellent enantioseparation properties involving a variety of guest compounds such as chiral alcohols, ketones, amines, sulfoxides, sulfoximines, and oxiranes.

Also among hydrogen-bonded network type of clathrate hosts there is a promising development leading to a class of helical tubulands of which [147318-38-5] (43) shows a prototypical structure (121). The building principle of clathrates formed by this family of diol hosts is a series of tight spiral spines of hydrogen bonds (Fig. 18b) (122). Host molecules radiate from and interconnect these spines such that a hexagonal arrangement of the spines enclose a channel which may contain trapped solvent molecules of a variety of functional group classes (hydrocarbons, aryl halides, ethers, esters, ketones, amines, nitriles, and sulfides). Variation of the diol molecular structure (eg, bridging of (43), equatorial orientation of the hydroxy groups) results in considerable modification of the channel topology and dimensions with altered inclusion properties (123).

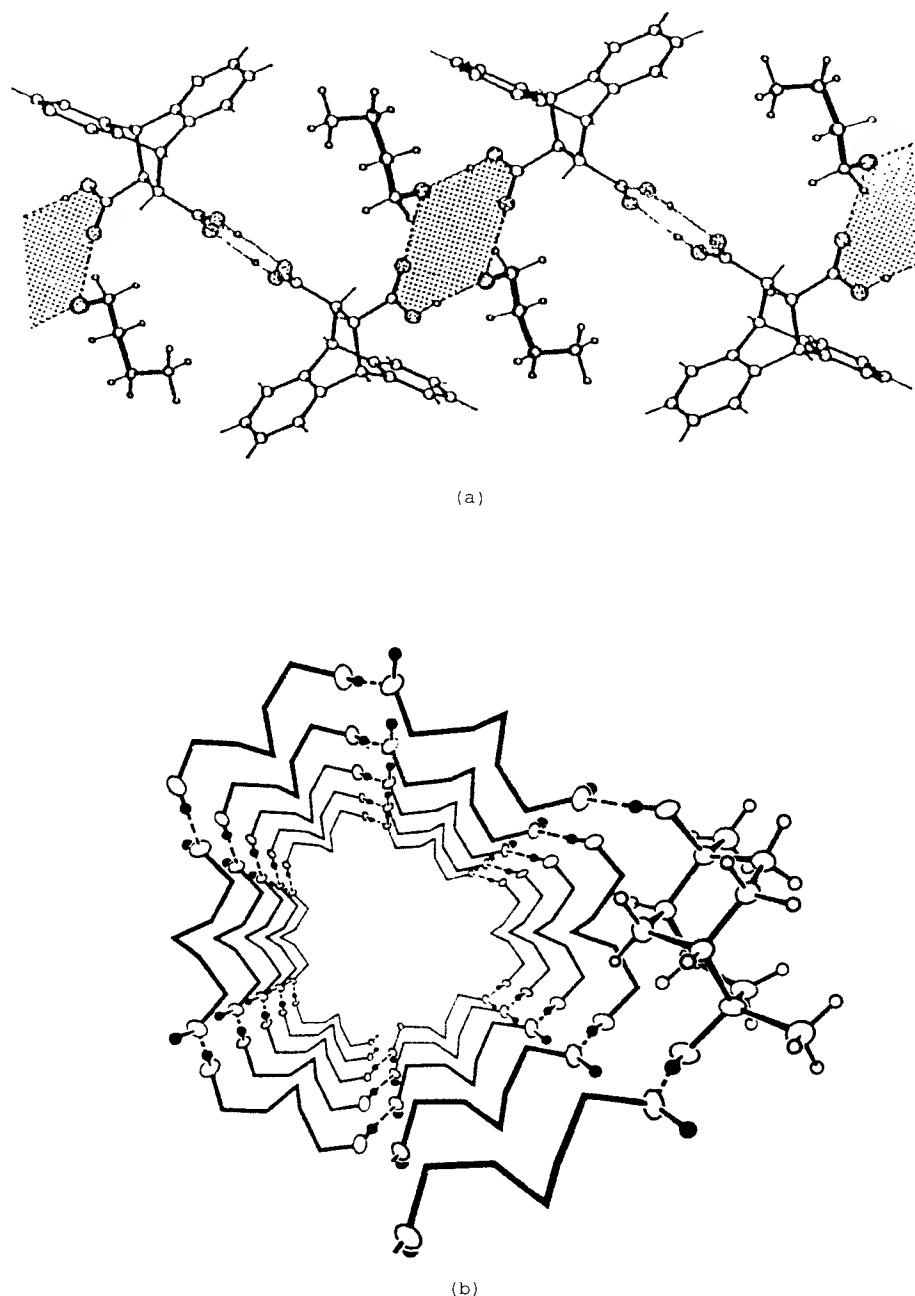


Fig. 18. Crystal structures of recent clathrate design: (a) coordinatoclathrate between host (39) (Fig. 17) and *n*-butanol (host-guest hydrogen bonding in the shaded area); (b) perspective view of the helical inclusion channel formed by diol host (43) (Fig. 17; all except one host molecule are represented diagrammatically) (111,122).

Preparation and Characterization of Inclusion Compounds

There are several ways to prepare inclusion compounds. In solution (25,44), they may simply be formed by dissolving together host and guest in a common solvent. This solvent should compete minimally with the guest or, preferably, the bulk solvent should be the guest compound. Inclusion formation in solution applies only for intramolecular cavity inclusions and complexes. Crystalline inclusion compounds (21) may be prepared by crystallization from the guest solvent or by cocrystallization of host and guest from an inert solvent. Solid inclusion compounds are also formed by direct exposure of the host to the vapor or liquid guest or, sometimes (124,125), by grinding solid host and guest together. Moreover, replacement of an included guest has been demonstrated in particular cases (99,100).

The formation of such materials may be monitored by several techniques. One of the most useful methods is ^1H - and ^{13}C -nmr spectroscopy where stable complexes in solution may give rise to characteristic shifts of signals relative to the uncomplexed species (43). Solution nmr spectroscopy has also been used to detect the presence of solid inclusion compound (after dissolution) and to determine composition (host:guest ratio) of the material. Infrared spectroscopy (126) and combustion analysis are further methods to study inclusion formation. For general screening purposes of solid inclusion structures, the x-ray powder diffraction method is suitable (123). However, if detailed structures are required, the single crystal x-ray diffraction method (127) has to be used.

Appropriate guest molecules are those that have a suitable size and shape to accommodate the host cavity and that complement the host cavity chemically. Host cavities lined only with apolar groups may prefer apolar guest entrapment and polar host cavities favor polar guests. Host cavity (cage) free diameters usually range between 0.4 and 0.8 nm. Host channels are inclined to accommodate linear guests; host cages have affinity to entrap spherical guests. Cage inclusions are generally more stable than channel or layer inclusions.

Stabilities of inclusion compounds span a wide range. Some are very stable at ambient conditions and require heating to considerable temperatures or treatment under high vacuum to cause decomposition. Others are only stable when in contact with mother liquor or excess guest solvent from which the inclusion compound was grown. A simple yet informative way for estimation of inclusion stabilities is to relate the decomposition point of the inclusion compound to the usual boiling point of the respective guest liquid (110). However, thermodynamic and kinetic properties including stabilities of inclusion compounds have also been studied on a more scientific base (44,128,129). Stability data of crown cation and uncharged molecule complexes of macrocyclic and oligocyclic hosts in solution are available in numerous cases (38,39).

Uses

Inclusion compounds open up a wide area of applications (1,2,17–28). An important aspect in this connection is the specific microenvironment created by the host enclosure of the guest which exerts an influence on the physical, spectroscopic, chemical, and other properties of the guest.

Retardation and Control. This influence may manifest itself in a reduced volatility and therefore lower possible storage and handling problems of a compound when included; toxic and hazardous substances become safer. As most inclusion compounds dissociate in a solvent or decompose under intensified conditions, it is easy to recover the guest. 18-crown-6 ((3) Fig. 3), for example, forms a solid inclusion complex with dimethyl sulfate that makes this dangerous compound easy to handle and to dose (130). The hydroquinone clathrate of krypton can be used as a way of handling radioactive krypton (^{85}Kr) turned into a solid, thus providing a safe and useful radioactive source (131). Moreover, odorous substances and flavoring compounds may be solidified via inclusion and reduced in their vapor pressure and volatility which makes controlled, retarded, and suppressed release possible. Cyclodextrin inclusions (Fig. 7) are most useful in this sense and aroused great value for many industries including cosmetology, food industry, laundering, pharmacy, agriculture, and others (57,132,133). For instance, cyclodextrins can be used to trap clinging and unpleasant odors (fish, garlic, cigarettes, alcohol). They can be used in the form of liquids, powders, tablets, chewing gums, toothpastes, sprays, and mouthwashes (133). Interesting examples of prolonged release obtained with cyclodextrins concern drugs (134), pesticides (135), and pheromones (136). An extreme form of molecular imprisonment and environmental shielding is found in carceplexes (see (17) Fig. 5) that can release guest molecules only by breaking covalent bonds of the container shell (46).

Enclosure also changes the redox properties of a compound, its color, and other physical properties (1,2). On this basis nonlinear optical materials, luminescence markers, controlled light switches, and other high-tech devices might be designed and prepared (15,17,137).

Shielding and Stabilization. Inclusion compounds may be used as sources and reservoirs of unstable species. The inner phases of inclusion compounds uniquely constrain guest movements, provide a medium for reactions, and shelter molecules that self-destruct in the bulk phase or transform and react under atmospheric conditions. Clathrate hosts have been shown to stabilize molecules in unusual conformations that can only be obtained in the host lattice (138) and to stabilize free radicals (139) and other reactive species (1) similar to the use of matrix isolation techniques. Inclusion compounds do, however, have the great advantage that they can be used over a relatively wide temperature range. Cyclobutadiene, pursued for over a century has been generated photochemically inside a carcerand container (see (17) Fig. 5) where it is protected from dimerization and from reactants by its surrounding shell (140).

Improved stability of included guest molecules leading to protection from environmental factors such as heat, light, and oxygen is also economical. In the food industry, spices and fruit flavors included and transformed into powder by cyclodextrins ((23), Fig. 6) exhibit good stability when they are heated during industrial food processing (133). So it is possible to use smaller amounts and such flavors last for a longer period. Similarly, oxidation of vitamins is considerably slowed down (141) and insecticides such as pyrethroids included in cyclodextrins show higher resistance to ultraviolet light and oxygen (142). Clathrates of fatty acids have been used to protect them from oxidation (88). Numerous dyes form stable inclusion complexes with cucurbituril ((24) Fig. 9) which leads to stabilization of such molecules (65) and has importance in the textile industry (143).

Solubilization and Activation. Compounds included in a host take solubility properties of the host shell and thus become more soluble when trapped in polar or apolar media, depending on the nature of the host. This leads to important uses in chemical synthesis known as the phase-transfer principle (144). Salt-type reagents (nucleophiles, bases, etc) including a metal ion that are only poorly soluble in organic solvents may be drastically enhanced in their solubility via complexation with crown compounds or cryptands (see Fig. 3) when used in solid–liquid and liquid–liquid (aqueous–organic solvent) phase-transfer systems (see CATALYSIS, PHASE TRANSFER). Complexation of the cation and transfer of the reagent into a lipophilic medium causes activation of poorly solvated anions (naked anions) that work miracles in reactivity (145–147). Very resistant esters can be saponified or problematic oxidation and substitution reactions can be performed using this method. The other way round, cyclophane type hosts (Fig. 5) (43,44) and cyclodextrins (Fig. 6) (56,57) enhance hydrosolubilities of organic molecules. Furthermore, increased compound solubility occupies an important place in pharmacy, cosmetics, or the food industry (139). For instance, sweetening agents which precipitate easily by cooling, no longer precipitate after addition of β -cyclodextrin ((23b), Fig. 6) (133); complexed flurbiprofen becomes more soluble in water (148).

Organized Media Effects. Another general reason for using host–guest inclusion chemistry in synthesis is controlled selectivity and artificial enzyme mimicry (149–151). Anisole included in α -cyclodextrin ((23a), Fig. 6) gives a considerably higher ratio of *p*-chlorination than under usual conditions (152). Cyclodextrins and cyclophane hosts (see Fig. 5) carrying catalytic groups show efficiency in hydrolysis or C–C coupling reactions (43,44,153). The same types of host also work as molecular reaction vessels via cooperative binding of reactants, eg, to perform an accelerated Diels–Alder reaction (152). The organizing principle of crystalline inclusion compounds make topochemistry (154) and crystal engineering (155) possible. Evidently, channel complexes (see Fig. 16) can act as templates for stereoregular inclusion polymerizations (156,157). Photoreactions of clathrates give rise to regio- and stereo-controlled products different from solution reaction (105). Enantioselective product formation via solid state reaction (158) can be performed in bulk from achiral material under absolute asymmetric conditions.

Separation and Retrieval. Separation of chemical species that differ in shape and size is another important field of application. All kinds of metal ions, anions, and uncharged molecules that match a specific cavity structure, thus forming a stable inclusion complex, make a distinction from nonmatching materials possible. Crown compounds and analogues (see Fig. 3) are efficient in separating individual metal ions such as alkali and alkaline-earth metal ions, and also transition metals if the host donor sites are suitable (159,160). They are useful for recovery of noble metals, for decontamination of wastewaters, and other concentration processes based on solvent extraction (161) and membrane separation techniques (162,163). Efficiency and convenience is high if the host is attached to an insoluble support (164). Hosts for molecular species are able to separate aromatic from aliphatic compounds (107), substituted from unsubstituted compounds (105), linear from branched isomers (5), positional isomers (165), diastereomers, and enantiomers (117,166–168). Racemate resolution using clathrates (eg, of (38), (41), and (42); Fig. 17) are noticed increasingly (105,107,118,169). Dyes, alkaloids, carbohydrates, nucleotides, or barbiturates are particular examples of compounds separated in bulk amounts (11,44).

Analytically, the inclusion phenomenon has been used in chromatography both for the separation of ions and molecules, in liquid and gas phase (1,79,170,171). Peralkylated cyclodextrins enjoy high popularity as the active component of hplc and gc stationary phases efficient in the optical separation of chiral compounds (57,172). Chromatographic isotope separations have also been shown to occur with the help of Werner clathrates and crown complexes (79,173).

Sensing. Crown compounds modified by responsible chromogenic groups (chromoionophores) (174) proved valuable tools for measuring metal ions (175,176) and even enantiomeric guest concentrations in solution (177). Ion selective electrodes based on crown compounds and podands (see Fig. 3) as the sensitive component (178,179) have broad analytical applications from industrial wastewater control to clinical bedside monitoring of blood (162). Moreover, cyclodextrins and other cavity hosts gain increasing influence as sensitive coatings of chemical sensor devices (180–183) for organic solvents and vapors (184,185). Quite recently, it appeared that clathrate forming hosts such as shown in Figure 17 are also useful in this field (186–188).

BIBLIOGRAPHY

"Clathration" in *ECT* 3rd ed., Vol. 6, pp. 179–189, by E. C. Makin, Monsanto Chemical Intermediates Co.

1. J. E. D. Davies and co-workers, *J. Incl. Phenom.* **1**, 3 (1983).
2. E. Weber, in E. Weber, ed., *Molecular Inclusion and Molecular Recognition - Clathrates I*, Springer, Berlin-Heidelberg, 1987, p. 1.
3. H. M. Powell, *J. Chem. Soc.*, 61 (1948).
4. H. M. Powell, in L. Mandelcorn, ed., *Non-stoichiometric Compounds*, Academic Press, Inc., New York, 1964, Chapt. 7.
5. F. Cramer, *Angew. Chem.* **64**, 437 (1952).
6. C. J. Pedersen, *J. Am. Chem. Soc.* **89**, 7017 (1967).
7. F. Vngtle and E. Weber, eds., *Host-Guest Complex Chemistry - Macrocycles*, Springer, Berlin-Heidelberg, 1985.
8. J. F. Stoddart, *Ann. Rep. Progr. Chem.* **85**, 353 (1988).
9. G. Van Binst, ed., *Design and Synthesis of Organic Molecules Based on Molecular Recognition*, Springer, New York, 1986.

10. H.-J. Schneider, *Angew. Chem.* **103**, 1419 (1991); *Angew. Chem. Int. Ed. Engl.* **30**, 1417 (1991).
11. A. D. Hamilton, in H. Dugas, ed., *Frontiers in Bioorganic Chemistry*, Vol. 2, Springer, Berlin, 1991, p. 115.
12. E. Fischer, *Ber. Dtsch. Chem. Ges.* **27**, 2985 (1894).
13. C. J. Pedersen, *Angew. Chem.* **100**, 1053 (1988); *Angew. Chem. Int. Ed. Engl.* **27**, 1021 (1988); C. J. Pedersen, *J. Incl. Phenom.* **6**, 337 (1988).
14. D. J. Cram, *Angew. Chem.* **100**, 1041 (1988); *Angew. Chem. Int. Ed. Engl.* **27**, 1009 (1988); D. J. Cram, *J. Incl. Phenom.* **6**, 397 (1988).
15. J.-M. Lehn, *Angew. Chem.* **100**, 91 (1988); *Angew. Chem. Int. Ed. Engl.* **27**, 89 (1988); J.-M. Lehn, *J. Incl. Phenom.* **6**, 351 (1988).
16. E. Weber and F. Vögtle, *Nachr. Chem. Techn. Lab.* **35**, 1149 (1987).
17. F. Vögtle, *Supramolecular Chemistry - An Introduction*, John Wiley & Sons, Inc., Chichester, U.K., 1991.
18. V. Balzani and De Cola, eds., *Supramolecular Chemistry*, Kluwer, Dordrecht, 1992.
19. G. W. Gokel, ed., *Advances in Supramolecular Chemistry*, Vol. 1, JAI-Press, Greenwich, Conn., 1990.
20. E. Weber, ed., *Supramolecular Chemistry I - Directed Synthesis and Molecular Recognition*, Springer, Berlin-Heidelberg, 1993.
21. J. L. Atwood, J. E. D. Davies, and D. D. MacNicol, eds., *Inclusion Compounds*, Vols. 1–3, Academic Press, Inc., London, 1984; Vols. 4–5, Oxford University Press, Oxford, 1991.
22. E. Weber, ed., *Molecular Inclusion and Molecular Recognition - Clathrates I and II*, Springer, Berlin-Heidelberg, 1987 and 1988.
23. E. Weber and H.-P. Josel, *J. Incl. Phenom.* **1**, 79 (1983).
24. S. Patai and Z. Rappoport, eds., *Crown Ethers and Analogs*, John Wiley & Sons, Inc., Chichester, U.K., 1989.
25. G. W. Gokel, *Crown Ethers and Cryptands, Monographs in Supramolecular Chemistry*, Vol. 3, The Royal Society of Chemistry, Cambridge, 1991.
26. M. Hiraoka, ed., *Crown Ethers and Analogous Compounds*, Elsevier, Amsterdam, 1992.
27. S. R. Cooper, ed., *Crown Compounds Toward Future Applications*, VCH, Weinheim, 1992.
28. B. Dietrich, P. Viout, and J.-M. Lehn, *Macrocyclic Chemistry*, VCH, Weinheim, 1993.
29. J.-M. Lehn, *Acc. Chem. Res.* **11**, 49 (1978).
30. F. Vögtle and E. Weber, *Angew. Chem.* **91**, 813 (1979); *Angew. Chem. Int. Ed. Engl.* **18**, 753 (1979).
31. E. Weber, in Ref. 24, p. 305.
32. G. W. Gokel, in Ref. 21, Vol. 4, p. 283.
33. F. Vögtle and E. Weber, *Angew. Chem.* **86**, 896 (1974); *Angew. Chem. Int. Ed. Engl.* **13**, 814 (1974).
34. D. J. Cram, *Angew. Chem.* **98**, 1041 (1986); *Angew. Chem. Int. Ed. Engl.* **25**, 1039 (1986).
35. J. L. Toner, in Ref. 24, p. 77.
36. J. F. Stoddart, *Top. Stereochem.* **17**, 207 (1987).
37. D. J. Cram and J. M. Cram, *Acc. Chem. Res.* **11**, 8 (1978).
38. Y. Inoue and G. W. Gokel, eds., *Cation Binding by Macrocycles*, Marcel Dekker, Inc., New York, 1990.
39. R. M. Izatt and co-workers, *Chem. Rev.* **92**, 1261 (1992).
40. R. Hilgenfeld and W. Saenger, *Top. Curr. Chem.* **101**, 1 (1982).
41. I. Goldberg, in Ref. 24, p. 359 and 399.
42. E. Weber, in R. M. Izatt and J. J. Christensen, eds., *Synthesis of Macrocycles - The Design of Selective Complexing Agents, Progress in Macrocyclic Chemistry*, Vol. 3, John Wiley & Sons, Inc., New York, 1987, p. 337.
43. F. Diederich, *Angew. Chem.* **100**, 372 (1988); *Angew. Chem. Int. Ed. Engl.* **27**, 362 (1988).
44. F. Diederich, *Cyclophanes, Monographs in Supramolecular Chemistry*, Vol. 2, The Royal Society of Chemistry, Cambridge, 1991.
45. F. Vögtle, *Cyclophane Chemistry*, John Wiley & Sons, Inc., Chichester, 1993.
46. D. J. Cram, *Nature* **356**, 29 (1992).
47. A. Collet, in Ref. 20, p. 103.
48. F. H. Kohnke, J. P. Mathias, and J. F. Stoddart, *Angew. Chem.* **101**, 1129 (1989); *Angew. Chem. Int. Ed. Engl.* **28**, 1103 (1989).
49. F. H. Kohnke, J. P. Mathias, and J. F. Stoddart, in Ref. 20, p. 1.
50. B. Odell and co-workers, *Angew. Chem.* **100**, 1605 (1988); *Angew. Chem. Int. Ed. Engl.* **27**, 1547 (1988).
51. J. F. Stoddart, in H.-J. Schneider and M. Dörr, eds., *Frontiers in Supramolecular Chemistry and Photochemistry*, VCH, Weinheim, 1991, p. 251.
52. C. D. Gutsche, *Calixarenes, Monographs in Supramolecular Chemistry*, Vol. 1, The Royal Society of Chemistry, Cambridge, 1989.
53. J. Vicens and V. Böhmer, eds., *Calixarenes - A Versatile Class of Macrocyclic Compounds*, Kluwer, Dordrecht, 1991.
54. A. Arduini and co-workers, in Ref. 18, p. 31.
55. L. C. Groenen and D. N. Reinhoudt, in Ref. 18, p. 51.
56. M. L. Bender and M. Komiyama, *Cyclodextrin Chemistry*, Springer, New York, 1978.
57. J. Szejtli, *Cyclodextrin Technology*, Kluwer, Dordrecht, 1988.
58. W. Saenger, in Ref. 21, Vol. 2, p. 231.
59. K. Harata, in Ref. 21, Vol. 5, p. 311.
60. R. J. Bergeron, in Ref. 21, Vol. 3, p. 391.
61. I. Tabushi, in Ref. 21, Vol. 3, p. 445.
62. R. Breslow, *Supramol. Chem.* **1**, 111 (1993).
63. H.-T. Moon and co-workers, *Supramol. Chem.* **1**, 327 (1993).
64. W. Saenger, *Naturwissenschaften* **71**, 31 (1984).
65. P. Cintas, *J. Incl. Phenom.* **17**, 205 (1994).
66. W. A. Freeman, W. L. Mock, and N. Y. Shih, *J. Am. Chem. Soc.* **103**, 7367 (1981).
67. H.-J. Buschmann and E. Schollmeyer, *J. Incl. Phenom.* **14**, 91 (1992).
68. A. Flinn and co-workers, *Angew. Chem.* **104**, 1550 (1992); *Angew. Chem. Int. Ed. Engl.* **31**, 1475 (1992).
69. J. Rebek, Jr., *Science* **235**, 1437 (1987).
70. J. Rebek, Jr., in Ref. 22, Vol. 149, p. 189.
71. S. C. Zimmerman, in Ref. 20, p. 71.
72. J.-M. Lehn, *Acc. Chem. Res.* **11**, 49 (1978).
73. F. P. Schmidtchen, *Top. Curr. Chem.* **132**, 101 (1986).
74. F. P. Schmidtchen, *Nachr. Chem. Tech. Lab.* **36**, 8 (1988).
75. H. E. Katz, in Ref. 21, Vol. 4, p. 391.
76. G. A. Jeffrey, in Ref. 21, Vol. 1, p. 135.
77. Y. A. Dyadin, I. V. Bondaryuk, and F. V. Zhurko, in Ref. 21, Vol. 5, p. 213.
78. E. D. Sloan, Jr., *Clathrate Hydrates of Natural Gases*, Marcel Dekker, New York, 1989.
79. G. Tsoucaris, in G. R. Desiraju, ed., *Organic Solid State Chemistry*, Elsevier, Amsterdam, 1987, p. 207.
80. G. A. Jeffrey and R. K. MacMillan, *Progr. Inorg. Chem.* **8**, 43 (1967).
81. G. A. Jeffrey and W. Saenger, *Hydrogen Bonding in Biological Structures*, Springer, Berlin-Heidelberg, 1991.
82. T. Iwamoto, in Ref. 21, Vol. 1, p. 29.
83. J. Lipkowski, in Ref. 21, Vol. 1, p. 59.
84. L. R. Nassimbeni and M. L. Niven, in International Union of Crystallography, eds., *Molecular Structure - Chemical Reactivity and Biological Activity*, Oxford University Press, Oxford, 1988.

85. K. A. Hofmann and F. A. Kospert, *Z. Anorg. Allg. Chem.* **15**, 204 (1897).
86. T. Iwamoto, in Ref. 21, Vol. 5, p. 177.
87. K. Takemoto and N. Sonoda, in Ref. 21, Vol. 2, p. 47.
88. S. G. Frank, *J. Pharm. Sci.* **64**, 1585 (1975).
89. F. Cramer, *Einschlüßverbindungen*, Springer, Berlin, 1954.
90. W. Schlenk, Jr., *Fortschr. Chem. Forsch.* **2**, 92 (1951); *Acta Crystallogr.* **5**, 224 (1952).
91. D. D. MacNicol, in Ref. 21, Vol. 2, p. 1.
92. J. F. Brown, Jr., *Sci. Am.* **207**, 82 (1962).
93. D. D. MacNicol, J. J. McKendrick, and D. R. Wilson, *Chem. Soc. Rev.* **7**, 65 (1978).
94. A. Gdaniec, B. T. Ibragimov, and S. A. Talipov, *J. Incl. Phenom.* **9**, 231 (1990) and previous papers of this series.
95. B. T. Ibragimov, S. A. Talipov, and P. M. Zorky, *Supramol. Chem.* **3**, 147 (1994).
96. E. Giglio, in Ref. 21, Vol. 2, p. 207.
97. M. Farina, in Ref. 21, Vol. 2, p. 69.
98. G. Allegra and co-workers, *J. Chem. Soc. (B)*, 1028 (1967).
99. H. R. Allcock, *Acc. Chem. Res.* **11**, 81 (1978).
100. H. R. Allcock, in Ref. 21, Vol. 1, p. 351.
101. A. Collet, in Ref. 21, Vol. 2, p. 97.
102. R. Gerdil, in Ref. 22, Vol. 140, p. 71.
103. W. D. Ollis and J. F. Stoddart, in Ref. 21, Vol. 2, p. 169.
104. D. D. MacNicol, in Ref. 21, Vol. 2, p. 123.
105. F. Toda, in Ref. 21, Vol. 4, p. 126.
106. H. Hart, L.-T. W. Lin, and D. L. Ward, *J. Am. Chem. Soc.* **106**, 4043 (1984).
107. E. Weber, in Ref. 21, Vol. 4, p. 188.
108. E. Weber, S. Finge, and I. Csörgő, *J. Org. Chem.* **56**, 7281 (1991).
109. I. Goldberg, in Ref. 21, Vol. 4, p. 406.
110. E. Weber and co-workers, *J. Am. Chem. Soc.* **106**, 3297 (1984).
111. E. Weber and M. Czugler, in Ref. 22, Vol. 149, p. 45.
112. E. Weber, *J. Mol. Graphics* **7**, 12 (1989).
113. E. Weber, N. Dürpinghaus, and I. Csörgő, *J. Chem. Soc., Perkin Trans. 2*, 2167 (1990).
114. E. Weber and co-workers, *J. Am. Chem. Soc.* **111**, 7866 (1989).
115. L. J. Barbour and co-workers, *Supramol. Chem.* **1**, 331 (1993).
116. I. Goldberg and co-workers, *J. Chem. Soc., Perkin Trans. 2*, 953 (1990).
117. F. Toda, in Ref. 22, Vol. 140, p. 43.
118. E. Weber, K. Skobridis, and C. Wimmer, *GIT Fachz. Lab.* **36**, 740 (1992).
119. F. Toda and K. Tanaka, *Tetrahedron Lett.* **29**, 551 (1988).
120. E. Weber and C. Wimmer, *Chirality* **5**, 331 (1993).
121. R. Bishop and I. G. Dance, in Ref. 21, Vol. 4, p. 1.
122. R. Bishop and I. G. Dance, in Ref. 22, Vol. 149, p. 137.
123. R. Bishop and co-workers, *Supramol. Chem.* **2**, 123 (1993).
124. F. Toda, K. Tanaka, and A. Sekikawa, *J. Chem. Soc., Chem. Commun.*, 279 (1987).
125. M. C. Etter and G. M. Frankenbach, *Materials* **1**, 10 (1989).
126. J. E. D. Davies, in Ref. 21, Vol. 3, p. 37.
127. G. D. Andreotti, in Ref. 21, Vol. 3, p. 129.
128. N. G. Parsonage and L. A. K. Staveley, in Ref. 21, Vol. 3, p. 1.
129. S. A. Bourne and co-workers, *J. Org. Chem.* **57**, 2438 (1992).
130. F. Vngtje and W. M. Möller, *Naturwissenschaften* **67**, 255 (1980).
131. D. J. Chleck and C. A. Ziegler, *Int. J. Appl. Radiat. Isotop.* **7**, 141 (1959).
132. J. Szejtli, in Ref. 21, Vol. 3, p. 331.
133. D. Duchkne and D. Wouessidjewe, *J. Coord. Chem.* **27**, 223 (1992).
134. J. Szejtli, in V. F. Smolen and L. A. Ball, eds., *Controlled Drug Bioavailability*, Vol. 3., John Wiley & Sons, Inc., New York, 1985, p. 365.
135. L. Szenté and J. Szejtli, *Acta Chim. Acad. Sci. Hung.* **107**, 196 (1981).
136. P. Kondilis and co-workers, in D. Duchkne, ed., *Minutes of the 5th Int. Symp. on Cyclodextrins*, Editions de Santü, Paris, 1990, p. 578.
137. G. A. Ozin, *Adv. Mater.* **4**, 612 (1992).
138. D. D. MacNicol and A. Murphy, *Tetrahedron Lett.* **22**, 1131 (1981).
139. T. Iwamoto, M. Kiyoki, and N. Matsuro, *Bull. Chem. Soc. Jpn.* **51**, 390 (1978).
140. D. J. Cram, M. E. Tanner, and R. Thomas, *Angew. Chem.* **103**, 1048 (1991); *Angew. Chem. Int. Ed. Engl.* **30**, 1024 (1991).
141. J. Szejtli and E. Bolla-Pusztai, *Störke* **32**, 386 (1980).
142. Y. Katsuda, H. Hirobe, and Y. Minamite, in J. Miyamoto, ed., *Pesticide Chemistry*, Vol. 2, Pergamon Press, Oxford, 1983, p. 223.
143. H.-J. Buschmann and co-workers, *Textilveredlung* **26**, 160 (1991).
144. E. V. Dehmlow and S. S. Dehmlow, *Phase Transfer Catalysis*, 3rd ed., VCH, Weinheim, 1993.
145. G. W. Gokel and H. D. Durst, *Synthesis*, 168 (1976).
146. E. Weber, *Kontakte (Merck)*, 38 (1983 1) and previous papers of this series.
147. C. L. Liotta, in Ref. 24, p. 59.
148. M. Otagiri and co-workers, *Acta Pharm. Suec.* **20**, 1 (1983).
149. B. S. Green, Y. Ashani, and D. Chipman, eds., *Chemical Approaches to Understanding Enzyme Catalysis: Biomimetic Chemistry and Transition State Analogs*, Elsevier, Amsterdam, 1982.
150. F. Diederich, *Chem. Uns. Zeit* **17**, 105 (1983).
151. J. F. Stoddart, in M. I. Page and A. Williams, eds., *Enzyme Mechanisms*, The Royal Society of Chemistry, Cambridge, 1987.
152. R. Breslow, in Ref. 21, Vol. 3, p. 473.
153. Y. Murakami, J. Kikuchi, and Y. Hisaeda, in Ref. 21, Vol. 4, p. 448.
154. G. R. Desiraju, ed., *Organic Solid State Chemistry*, Elsevier, Amsterdam, 1987.
155. G. Desiraju, *Crystal Engineering, Materials Science Monographs*, Vol. 54, Elsevier, Amsterdam, 1989.
156. M. Farina, in Ref. 21, Vol. 3, p. 297.
157. M. Miyata, in C. M. Paleos, ed., *Polymerization in Organized Media*, Gordon and Breach, Philadelphia, Pa., 1992, p. 327.
158. F. Toda, in Ref. 22, Vol. 149, p. 211.
159. Y. Takeda, in F. Vngtje and E. Weber, eds., *Host Guest Complex: Chemistry III*, Springer, Berlin, 1984, p. 1.
160. T. M. Fyles, in H. Dugas, ed., *Bioorganic Chemistry Frontiers*, Vol. 1, Springer, Berlin, 1990, p. 71.
161. D. H. Logsdail and M. J. Slater, eds., *Solvent Extraction in the Process Industries*, Vols. 1–3, Elsevier, New York, 1993.

162. E. Weber, *Kontakte (Merk)*, 26 (1984 1) and previous papers of this series.
163. J. D. Lamb and J. J. Christensen, in Ref. 21, Vol. 3, p. 571.
164. J. Smid and R. Sinta, in Ref. 159, p. 105.
165. K. Pollmer, *Z. Chem.* **19**, 81 (1979).
166. D. Worsch and F. Vngtle, in Ref. 22, Vol. 140, p. 21.
167. B. S. Green, M. Knossow, and G. Tsoucaris, in Ref. 21, Vol. 3, p. 263.
168. T. H. Webb and C. S. Wilcox, *Chem. Soc. Rev.*, 383 (1993).
169. E. Weber and co-workers, *J. Chem. Soc., Chem. Commun.*, 733 (1992); see also *Chem. Ind. (London)*, 364 (1992).
170. E. Blasius and K.-P. Janzen, in Ref. 7, p. 189.
171. D. Sybilska and E. Smolkov6-Keulemansov6, in Ref. 21, Vol. 3, p. 173.
172. W. A. Kunig, *Kontakte (Merk)*, 3 (1990 2).
173. N. O. Smith, in Ref. 21, Vol. 3, p. 257.
174. H.-G. Lühr and F. Vngtle, *Acc. Chem. Res.* **18**, 65 (1985).
175. S. Shinkai and O. Manabe, in Ref. 7, p. 245.
176. M. Takagi, in Ref. 38, p. 465.
177. S. Misumi, in Ref. 20, p. 163.
178. D. Amman and co-workers, *Ion-Selective Electrode Rev.* **5**, 3 (1983).
179. K. Kimura and T. Shono, in Ref. 38, p. 429.
180. J. C. Lockhart, in Ref. 21, Vol. 5, p. 345.
181. T. E. Edmonds, ed., *Chemical Sensors*, Blackie, Glasgow, 1988.
182. D. N. Reinhoudt and E. J. R. Sudhmlter, *Adv. Mater.* **2**, 23 (1990).
183. K. Cammann and co-workers, *Angew. Chem.* **103**, 519 (1991); *Angew. Chem. Int. Ed. Engl.* **30**, 516 (1991).
184. F. L. Dickert and A. Haunschild, *Adv. Mater.* **5**, 887 (1993).
185. G. Gauglitz and co-workers, *Spektrum der Wissenschaft*, 92 (Jan. 1994).
186. A. Ehlen and co-workers, *Angew. Chem.* **105**, 116 (1993); *Angew. Chem. Int. Ed. Engl.* **32**, 110 (1993); see also *Chem. Ind. (London)*, 333 (1993).
187. K. Buhlmann and co-workers, *Fresenius J. Anal. Chem.* **348**, 549 (1994).
188. J. Reinbold and co-workers, *Sensors and Actuators B* **18–19**, 77 (1994).

Edwin Weber
Technische Universität Bergakademie Freiberg

INDANTHRENE, INDANTHRENE DYES.

See DYES, ANTHRAQUINONE

INDENE.

See HYDROCARBON RESINS

INDICATORS.

See HYDROGEN ION ACTIVITY

INDIGOID DYES.

See DYES, NATURAL

INDIUM AND INDIUM COMPOUNDS

Indium

Indium [7440-74-6], an element of Group 13 (IIIA) of the Periodic Table, occurs between gallium and thallium. It is a soft, lustrous, silver-white metal, highly malleable and ductile, having a face-centered tetragonal crystalline structure, $a = 0.32512\text{ nm}$, $c = 0.49467\text{ nm}$ at 26°C. Indium was discovered in 1863 in sphaleritic ore in Germany, and named for the indigo blue spectral lines that led to its identification.

The abundance of indium in the earth's crust is probably about 0.1 ppm, similar to that of silver. It is found in trace amounts in many minerals, particularly in the sulfide ores of zinc and to a lesser extent in association with sulfides of copper, tin, and lead. Indium follows zinc through flotation concentration, and commercial recovery of the metal is achieved by treating residues, flue dusts, slags, and metallic intermediates in zinc smelting and associated lead (qv) and copper (qv) smelting (see METALLURGY, EXTRACTIVE; ZINC AND ZINC ALLOYS).

Properties. Table 1 lists many of the physical, thermal, mechanical, and electrical properties of indium. The highly plastic nature of indium, which is its most notable feature, results from deformation from mechanical twinning. Indium retains this plasticity at cryogenic temperatures. Indium does not work-harden, can endure considerable deformation through compression, cold-welds easily, and has a distinctive cry on bending as does tin.

Table 1. Physical Properties of Indium

Property	Value
atomic weight	114.82
<i>atomic number</i>	49
melting point, °C	156.61
boiling point, °C	2080
latent heat of fusion, kJ kg ^a	28.47
latent heat of vaporization at bp, kJ kg ^a	1959.42
specific heat at 25°C, kJ(kg·K) ^a	0.233
coefficient of linear expansion, 0–100°C, $\times 10^6\text{ }^\circ\text{C}^{-1}$	24.8
electrical resistivity, $\Omega\cdot\text{m}$	
at 3.38 K	superconducting
20°C	84
154°C	291
181°C	301
222°C	319
280°C	348
electrode oxidation–reduction potential, ^b V	0.38
density, kg m ^{−3}	
at 20°C	7.300
164°C	7.026
volume increase on melting, %	2.5
thermal conductivity at 0°C, W (m·K)	83.7
surface tension at temperature, T in K between mp and bp, mN m(dyn cm)	602–0.1 T
relative isotopic abundance, wt %	
¹¹³ In	4.23
¹¹⁵ In	95.77

vapor pressure, <i>p</i> , at temperature, <i>T</i> in K between mp and bp, kPa ^c	log ₁₀ <i>p</i>	9.835	12,860	<i>T</i>	0.7 log ₁₀ <i>T</i>
thermal neutron cross section at 2200 m s, × 10 ⁻²⁸ m ² atom ^d					
absorption					190 ± 10
scattering					2.2 ± 0.5
Brinell hardness number, HB					0.9
tensile strength, MPa ^c					
at 295 K					1.6
76 K					15.0
4 K					31.9
elongation, %					22
modulus of elasticity, GPa ^c					12.74
Poisson's ratio at 20°C					0.4498

^a To convert J to cal, divide by 4.184.
^b Versus the hydrogen electrode at 0.0 V.
^c To convert kPa to atm, divide by 101.3; to convert MPa to atm, divide by 0.101; to convert GPa to atm, divide by 1.01 × 10⁻⁴.
^d To convert m² to barn, multiply by 10²⁸.

Indium metal is not oxidized by air at ordinary temperature, but at red heat burns to form the trioxide In₂O₃. On heating, indium also reacts directly with the metalloids, ie, arsenic, antimony, selenium, and tellurium, and with halogens, sulfur, and phosphorus. Indium dissolves in mineral acids and amalgamates with mercury but is minimally affected by alkalies, boiling water, and most organic acids. The metal surface passivates easily. Indium usually exhibits a valence of +3, but may also have valencies of +2 and +1. In general, the chemistry of indium in its trivalent compounds stems from nonionic or covalent bonding characteristics. Indium is electroplated easily from a variety of baths including the cyanide (1), sulfate (2), fluoborate (3), and sulfamate (4) (see ELECTROPLATING).

Production. Indium is recovered from fumes, dusts, slags, residues, and alloys from zinc or lead–zinc smelting. The source material itself, a reduction bullion, flue dust, or electrolytic slime intermediate, is leached with sulfuric or hydrochloric acid, the solutions are concentrated, if necessary, and crude indium is recovered as 99+ % metal. This impure indium is then refined to 99.99⁰ o, 99.999⁰ o, 99.9999⁰ o, or higher grades by a variety of classical chemical and electrochemical processes.

Economic Aspects. Production of indium has been reported from Belgium, Canada, China, France, Germany, Italy, Japan, the Netherlands, Peru, the United Kingdom, and the United States, as well as countries in the CIS (the former Soviet Union) (5).

Production reported from Belgium, the Netherlands, and the United Kingdom is thought to be from imported concentrates, residues, and scraps. Production in the CIS and the People's Republic of China is difficult to confirm. Many producers do not report indium production. The U.S. Bureau of Mines estimates world production in 1992 was 140 metric tons (4,500,000 troy oz), a 128⁰ o increase over the 1982 level, reflecting a significant increase in consumption. Research and product development programs indicate that a strong growth rate should continue for some time. Increases in indium production are expected to be easily accomplished.

Indium prices vary by purity level according to the number of mines. For 99.97 or 99.99⁰ pure indium *Metal Bulletin* magazine's free-market prices ranged from \$160 to \$190 /kg in 1993. Prices have trended downward since 1988. This, coupled with the fact that consumption increased during this period, is evidence that supply is ahead of demand. The most notable production increases have occurred at Indium Corp. of America, Nippon Mining Co. Ltd., Cominco Ltd., and Metaleurop S.A. The largest increase was made by Indium Corp. A 30-t increase was announced in 1988; in 1993 Indium Corp. announced plans to increase production by another 30 t to be phased in with demand, bringing the company's actual and planned increases to 60 metric tons over the 1988 level.

Analysis. Indium can be detected to 0.01 ppm by spectroscopic analysis, using its characteristic lines in the indigo blue region, at wavelengths 4511.36, 4101.76, 3256.09, and 3093.36 nm. Procedures for the quantitative determination of indium in ores, compounds, alloys, and for the analysis of impurities in indium metal are covered thoroughly in the literature (6).

Safety and Handling. Physiologically indium is a nonessential element (see MINERAL NUTRIENTS). It is classified as toxic, but there have been no reported cases of systemic effects in human exposure to indium. The threshold limit value of the ACGIH is 0.1 mg /m³. The primary toxic effects of ionic indium are on the kidneys. Indium has no demonstrated irritating effects on skin, but there may be effects on the respiratory system. Absorption through digestion is about 0.5⁰ o of intake and through respiration about 5⁰ o. There have been no reports of accidental or industrial poisoning, indicating that normal safety and containment measures are adequate.

Alloys. Indium alloys with a wide range of metals, and many binary and ternary systems have been studied extensively (7). Indium generally increases the strength, corrosion resistance, and hardness of a system to which it is added. Low melting point alloys of indium with lead, tin, bismuth, and cadmium having melting points as low as 47°C are used in surgical casts, patternmaking, lens blocking, turbine blade machining, fire door safety links, and sprinkler heads (see HIGH TEMPERATURE ALLOYS). The alloy of In 15⁰ o–Ag 85⁰ o–Cd 5⁰ o is used in control rods for nuclear reactors (qv). Tin–indium, lead–tin–indium, and lead–silver–indium solder alloys have melting points in the range of 100–300°C. These offer good corrosion resistance to alkalies, as well as improved resistance to thermal fatigue compared to conventional lead–tin solders. Several high indium alloys wet glass, quartz, and many ceramics. Indium also finds use in specialty brazing alloys with copper, silver, and gold. It is also a constituent of many dental alloys where it serves as an oxygen scavenger and hardener.

Uses. Indium's first commercial use was in the production of dental alloys (see DENTAL MATERIALS), but its first significant use was in the production of bearings for heavy-duty and high speed service (see BEARING MATERIALS). The advent of jet engines has reduced this use, but indium is still used in high performance engines.

The solder and alloy market, including low melting or fusible alloys, is a principal user of indium (see SOLDERS AND BRAZING ALLOYS). The addition of indium results in unique properties of solders such as improved corrosion and fatigue resistance, increased hardness, and compatibility with gold substrates. To facilitate use in various applications, indium and its alloys can be easily fabricated into wire, ribbon, foil, spheres, preforms, solder paste, and powder.

The supplanting of germanium-based semiconductor devices by silicon devices has almost eliminated the use of indium in the related alloy junction (see SEMICONDUCTORS). Indium, however, is finding increased use in III–V compound semiconductors such as indium phosphide [22398-80-7] for laser diodes used in fiber optic communication systems (see ELECTRONIC MATERIALS; FIBER OPTICS; LIGHT GENERATION). Other important indium-containing semiconductors include indium arsenide [1303-11-3], indium antimonide [1312-41-0], and copper–indium–diselenide [12018-95-0].

Applications for electroplated indium coatings include indium bump bonding for silicon semiconductor die attachment to packaging substrates and miscellaneous applications where the physical or chemical properties of indium metal are desired as a plated deposit.

Electrically conductive films of clear indium oxide (doped with tin oxide) on glass (qv) or plastic are finding use in many applications (see THIN FILMS). These include conductive electrodes for liquid crystal and other flat panel displays, as well as windshield defoggers and deicers. The combination of the high transparency to visible light and electrical conductivity make these coatings well suited for these applications. Because these coatings also reflect infrared radiation while passing visible light, they are finding increasing use as low energy (low E) coatings on energy efficient windows for residential and commercial buildings. Also these coatings are used on the internal surfaces of low pressure sodium vapor lamps to reflect heat inward, resulting in higher operating temperatures and improved lamp efficiency.

The softness and ductility of indium make it an excellent material for sealing gaskets. Indium easily deforms under pressure, filling imperfections of the mating surfaces being joined. Indium retains most of its softness and plasticity to temperatures approaching absolute zero, making it usable in

cryogenic environments (see CRYOGENICS). Indium is also used in temperature standards.

Indium chemicals and electroplated metal deposits are replacing mercury (qv) in the manufacture of alkaline batteries (qv). Indium, like mercury, functions to reduce outgassing within the battery and promotes the uniform corrosion of the anode and cathode while the battery is under electrical load. Indium inorganic chemicals also find use as catalysts in various chemical processes.

Indium Compounds

The usual valence of indium is three, although monovalent and bivalent compounds of indium with oxygen, halogens, and Group 15 (VA) and 16 (VIA) elements are well known. The lower valence compounds tend to disproportionate into the trivalent compound and indium metal; the trivalent compounds are stable.

Halides. Indium trichloride [10025-83-8], InCl₃, can be made by heating indium in excess chlorine or by chlorinating lower chlorides. It is a white crystalline solid, deliquescent, soluble in water, and has a high vapor pressure. InCl₃ forms chloroindates, double salts with chlorides of alkali metals, and organic bases.

Indium dichloride [13465-11-7], InCl₂, made by heating indium in hydrogen chloride or by reduction of InCl₃ in H₂ HCl, forms colorless crystals. Indium monochloride [13465-10-6], InCl, can be formed by passing InCl₃ vapor over heating indium.

Mono-, di-, and trivalent bromides and iodides may be made by methods similar to the chlorides. The lower valence salts also disproportionate in water. Indium trifluoride [7783-52-0], InF₃, is sparingly soluble in water. It forms an ammonium double salt, 3NH₄F · InF₃ [15273-84-4], which decomposes on heating to indium nitride [25617-98-5], InN.

Oxides. Three oxides of indium are known: indium trioxide, also known as indium sesquioxide [1312-43-2], In₂O₃; indium suboxide [12030-22-7], In₂O; and indium monoxide [12136-26-4], InO. There are many compound oxides of indium with other metals, the most important of which is that with tin, commonly known as ITO. Indium trioxide is formed by burning indium in air or by calcining other indium salts. It is pale yellow when cold and red-brown when hot. Both crystalline and amorphous forms exist. The noncrystalline form dissolves easily in acids but the crystalline form is resistant to acid dissolution. In₂O₃ is easily reduced to metal by hydrogen and other reducing gases at elevated temperatures. Reduction under carefully controlled conditions gives the lower oxides.

Sulfates. Indium metal and its oxides dissolve in warm sulfuric acid to give a solution of the trisulfate [13464-82-9], In₂(SO₄)₃. It is a white, crystalline, deliquescent solid, readily soluble in water that forms double salts with alkali sulfates and some organic substituted ammonium bases. Concentration of the acidified trisulfate solution produces indium acid sulfate crystal [57344-73-7], In(HSO₄)₂, and other reaction conditions give basic sulfates.

Sulfides. The main sulfide of indium is In₂S₃ [12030-24-9], which can be prepared by heating the metal with sulfur or by precipitation from weak acid solutions of indium salts by H₂S. Precipitated In₂S₃ varies in color from yellow to red-brown, and in crystal size depending on formation conditions. It dissolves in acids and sodium sulfide solution. Other reported sulfides of indium are InS [12030-14-7], a red-brown solid; In₂S [12196-52-0]; and In₄S₅ [12142-00-5].

Other Salts. Indium nitrate trihydrate [13770-61-1], In(NO₃)₃ · 3H₂O, is a soluble salt prepared by dissolution of the metal or oxide in nitric acid. Indium phosphate [14693-82-4], InPO₄, is precipitated by adding phosphate ions to a solution of an indium salt. It is soluble in water.

Organic Compounds. The best known organic indium compounds are the alkyls, trimethyl indium [3385-78-2], In(CH₃)₃; triethyl indium [923-34-2], In(C₂H₅)₃; and triphenyl indium [3958-47-2], In(C₆H₅)₃, which may be made by treating indium with the corresponding mercury alkyls, or by the Grignard reaction (qv). The first two, when pure, ignite spontaneously in air. Indium oxalate [38051-36-4], In₂(C₂O₄)₃, and indium methoxide [40521-21-9], In(OCH₃)₃, are also known.

Intermetallic and Semiconducting Compounds. Indium forms intermetallic compounds with a great many metals and combinations of metals including alkali metals, magnesium, the iron group, rare earths, and precious metals such as the platinum group. Carbon-free indium-based fullerenes having formulas Na₉ In₉₇M₂, where M is Ni, Pd, or Pt, have been isolated (8).

Indium also combines with nonmetallic elements and with metalloids such as N, P, Sb, As, Te, and Se. Many of the latter compounds are semiconducting as are the oxide and sulfide. Indium antimonide [1312-41-0], InSb; indium arsenide [1303-11-3], InAs; and indium phosphide [22398-80-7], InP, are the principal semiconducting compounds. These are all prepared by direct combination of the highly purified elements at elevated temperature under controlled conditions.

BIBLIOGRAPHY

"Indium" in *ECT* 1st ed., Vol. 7, pp. 834–839, by A. A. Smith, Jr., American Smelting and Refining Co.; "Indium and Indium Compounds" in *ECT* 2nd ed., Vol. 11, pp. 581–584, by A. A. Smith, Jr., American Smelting and Refinery Co.; in *ECT* 3rd ed., Vol. 12, pp. 207–212, by E. F. Milner, Cominco Ltd. and C. E. T. White, Indium Corp. of America.

- 1. **U.S. Pat. 1,965,251** (July 3, 1934), W. S. Murray and D. Gray (to Oneida Community Ltd.).
- 2. H. B. Linford, *Trans. Electrochem. Soc.* **79**, 443 (1941).
- 3. **U.S. Pat. 2,409,983** (Oct. 22, 1946), W. M. Martz (to General Motors Corp.).
- 4. **U.S. Pat. 2,458,839** (Jan. 11, 1949), J. R. Dyer, Jr. and T. J. Rowan (to Indium Corp. of America).
- 5. *World Minor Metals Survey*, 2nd ed., Metal Bulletin, London, 1981.
- 6. J. M. Ramaradhya, R. C. Bell, S. Brownlow, and C. J. Mitchell, in S. D. Snell and L. S. Eltre, eds., *Encyclopedia of Industrial Chemical Analysis*, Vol. 19, John Wiley & Sons, Inc., New York, 1971, p. 518.
- 7. C. E. T. White and H. Okamoto, eds., *Phase Diagrams of Indium Alloys and their Engineering Applications*, ASM International, Materials Park, Ohio, 1992.
- 8. S. C. Sevov and T. C. Corbett, *Science* **262**, 880–883 (1993).

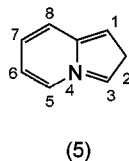
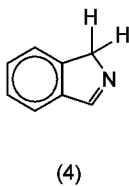
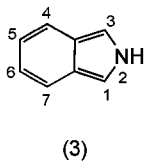
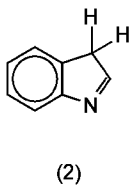
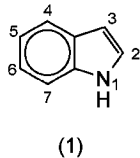
General References

M. T. Ludwick, *Indium*, 2nd ed., Indium Corp. of America, Utica, N.Y., 1959; a comprehensive bibliography.
L. G. Stevens and C. E. T. White, "Indium and Bismuth," in *Metals Handbook*, Vol. 2, 10th ed., ASM International, Materials Park, Ohio, 1990.

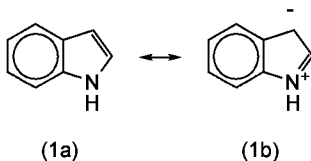
James A. Slattery
Indium Corporation of America

INDOLE

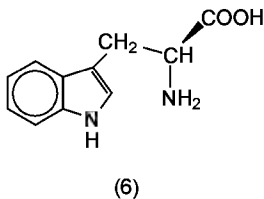
Indole is a heteroaromatic compound consisting of a fused benzene and pyrrole ring, specifically benzo[*b*]pyrrole. The systematic name, 1*H*-indole (**1**) distinguishes it from the less stable tautomer 3*H*-indole [271-26-1] (**2**). 1*H*-Indole [120-72-9] is also more stable than the isomeric benzo[*c*] pyrrole, which is called isoindole, (2*H*, (**3**) and 1*H* (**4**)). A third isomer benzo[*a*]pyrrole is a stable compound called indolizidine [274-40-8] (**5**).



Indole is planar with 10 π -electrons in a completely conjugated system. The ring is classified as a π -excessive heteroaromatic compound because of the electron-donating character of the pyrrole-type nitrogen atom. The π -system is relatively electron-rich, particularly at C-3, as represented by resonance structure (**1b**).



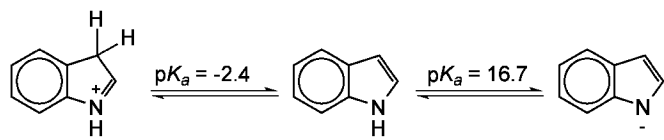
The indole ring is incorporated into the structure of the amino acid tryptophan [6912-86-3] (**6**) and occurs in proteins and in a wide variety of plant and animal metabolites. Much of the interest in the chemistry of indole is the result of efforts to understand the biological activity of indole derivatives in order to develop pharmaceutical applications.



Properties

Indole is a colorless solid, mp 52–54°C, which is readily soluble in most organic solvents but sparingly soluble in water. Indole has a musty odor which is very persistent and its derivatives have some applications in the formulation of fragrances.

Indole is a neutral compound but can be protonated or deprotonated under strongly acidic or basic conditions, respectively. The pK_a of the conjugate acid is about 2.4; that of the neutral compound is about 16.7 (1).



Crystal structure data are available for an indole–trinitrobenzene complex (2) and for the lithium and sodium salts in the presence of polyamine ligands (3). The crystal structure of indole itself is evidently disordered (4). Table 1 gives the 1H and ^{13}C -nmr assignments in $CDCl_3$ (5). ^{13}C -nmr assignments have been tabulated for many other indole derivatives (6).

Table 1. 1H -nmr Chemical Shifts for Indole and ^{13}C -nmr Chemical Shifts^a for Indole

Ring position	1H	^{13}C
1	7.73	
2	7.00	124.2
3	6.51	102.4
3a		127.8
4	7.64	120.7
5	7.11	119.8
6	7.18	121.9
7	7.24	111.1
7a		135.7

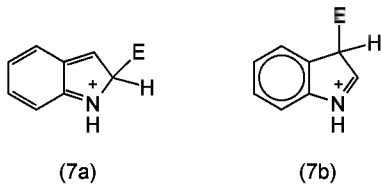
^a In $CDCl_3$.

The industrial source of indole has been isolation from coal-tar distillate (7). Several patents for the manufacture of indole have been issued with aniline and ethylene glycol (8), aniline and ethylene oxide (9), 2-ethylaniline (10), and *N*-ethylaniline (11) as the starting materials.

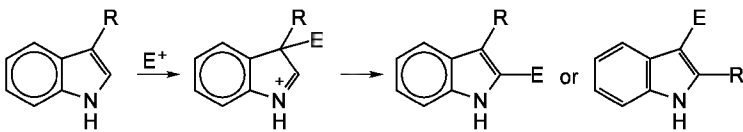
Reactivity

Indole is a heterocyclic analogue of naphthalene. The basic reactivity patterns of indole can be understood as resulting from the fusion of an electron-rich pyrrole ring with a benzene ring.

Electrophilic Aromatic Substitution. The π -excessive character of the pyrrole ring makes the indole ring susceptible to electrophilic attack. The reactivity is greater at the 3-position than at the 2-position. This reactivity pattern is suggested both by electron density distributions calculated by molecular orbital methods and by the relative energies of the intermediates for electrophilic substitution, as represented by the protonated structures (7a) and (7b). Structure (7b) is more favorable than (7a) because it retains the benzenoid character of the carbocyclic ring (12).



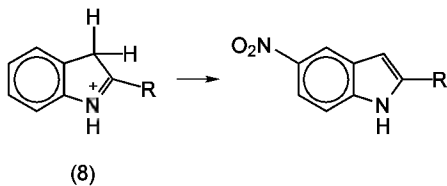
This basic reactivity pattern is not greatly affected by the presence of a 1- or 2- substituent, although electron-attracting substituents do diminish the reactivity. The pattern for substitution in 3-substituted indoles can be complicated by the fact that the electrophile may preferentially attack the 3-position, even when it is already substituted. When this is the case, migration of either the new or the original substituent to C-2 may occur.



Many of the common electrophilic aromatic substitution reactions can be conducted on indole. Complications normally arise either because of excessive reactivity or the relative instability of the substitution product. This is the case with halogenation.

Halogenation. 3-Chloroindole can be obtained by chlorination with either hypochlorite ion or with sulfuryl chloride. In the former case the reaction proceeds through a 1-chloroindole intermediate (13). 3-Chloroindole [16863-96-0] is quite unstable to acidic aqueous solution, in which it is hydrolyzed to oxindole. 3-Bromoindole [1484-27-1] has been obtained from indole using pyridinium tribromide as the source of electrophilic bromine. Indole reacts with iodine to give 3-iodoindole [26340-47-6]. Both the 3-bromo and 3-iodo compounds are susceptible to hydrolysis in acid but are relatively stable in base.

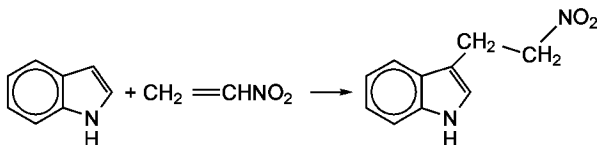
Nitration. Because nitration frequently generates nitrogen oxides which can participate in oxidative transformations, the nitration of indole itself is a complex reaction. In strongly acidic media, the nitration of 2-substituted indoles can proceed through the conjugate acid (8). Because the aromatic system is thereby transformed to an azastylene, the 5-position is the primary site of reaction.



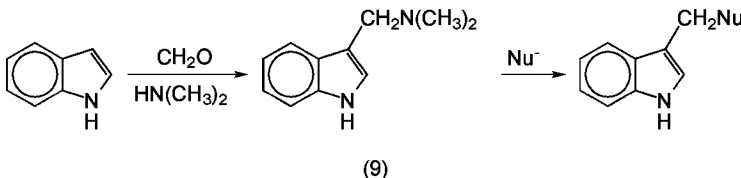
C-Alkylation. By choice of reaction conditions it is possible to favor alkylation of indole and substituted indoles at either the 1- or the 3-position. Reaction at the 3-position (*C*-alkylation) is favored by conditions which make the transition intermediate more like that in an electrophilic aromatic substitution. Tight metal coordination at nitrogen, as in the magnesium salt, promotes alkylation at C-3. Good yields are most likely to be obtained with highly reactive alkylating agents such as allylic and benzylic systems under conditions in which the alkylating agent assumes carbocationic

character.

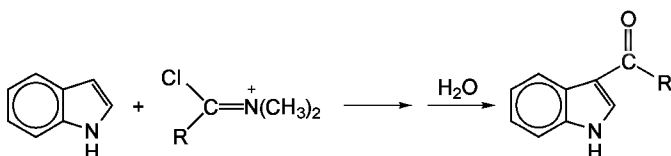
Unsaturated compounds activated by electron-attracting groups can also effect alkylation at the 3-position. An important example is the reaction with nitroethylene to form 3-(2-nitroethyl)indole [31731-23-4] (14).



Another useful reagent for the 3-alkylation of indole is the *N,N*-dimethylformaldiminium ion, which forms the useful intermediate gramine [87-52-5] (9). The C-3 substituent can subsequently be modified by displacement of the dimethylamino group by a nucleophile. Alternatively, gramine can be converted to its quaternary salt prior to substitution. A variety of carbanions can function as the nucleophile.

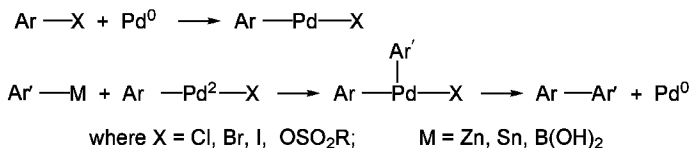


Acylation. Acylation is the most reliable means of introducing a 3-substituent on the indole ring. Because 3-acyl substituents can be easily reduced to 3-alkyl groups, a two-step acylation–reduction sequence is often an attractive alternative to direct 3-alkylation. Several kinds of conditions have been employed for acylation. Very reactive acyl halides, such as oxalyl chloride, can effect substitution directly without any catalyst. Normal acid chlorides are usually allowed to react with the magnesium (15) or zinc (16) salts. The Vilsmeier-Haack conditions involving an amide and phosphorus oxychloride, in which a chloroiminium ion is the active electrophile, frequently give excellent yields of 3-acylindoles.

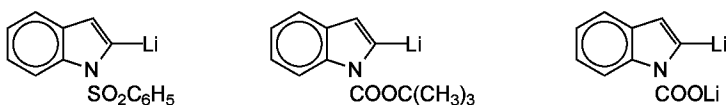


N-Alkylation. 1-Substitution is favored when the indole ring is deprotonated and the reaction medium promotes the nucleophilicity of the resulting indole anion. Conditions which typically result in *N*-alkylation are generation of the sodium salt by sodium amide in liquid ammonia, use of sodium hydride or a similar strong base in *N,N*-dimethylformamide or dimethyl sulfoxide, or the use of phase-transfer conditions.

Arylation. Arylation is normally accomplished through some substituted intermediate, rather than directly on indole. The only direct means of arylation, via radical substitution, is not selective enough to be useful for synthetic purposes. Palladium-catalyzed cross-coupling reactions have become the preferred means for arylation of indoles, as well as other heteroaromatic rings (17). Cross-coupling requires one nucleophilic component, typically an aryltin or arylzinc reagent. Indoleboronic acids can also serve as the nucleophilic component (18). The second component is an aryl halide or aryl triflate which can undergo an oxidative addition reaction with the palladium catalyst. Arylation occurs by addition of the nucleophilic component to the palladium intermediate and regenerates the active palladium species. Reactions have been reported in which the indole is either the nucleophilic or the oxidative reactant, but the former is the more common case.



Lithiation and Subsequent Transformations. Lithiation is the most general means of introducing a 2-substituent on the indole ring. Three intermediates have been used most frequently in this context. These are 1-phenylsulfonylindole (19), 1-*t*-butoxycarbonylindole (20), and lithium indole-1-carboxylate (21).



Each of these intermediates can be lithiated in the 2-position in good yield. The reactivity toward lithiation is due to the inductive effect of the nitrogen atom and coordination by oxygen from the *N*-substituent. A wide variety of electrophiles can then carry out substitution at the 2-position. Lithiation at other positions on the ring can be achieved by halogen–metal exchange; 3-lithio and 5-lithioindoles have also been used as reactive intermediates.

Oxidation. As a π -excessive heterocycle, indole is susceptible to oxidation; a variety of oxidation intermediates and products have been observed. With oxygen as the oxidant, the key intermediate is normally a 3-hydroperoxy-3*H*-indole. These intermediates are observable for 2,3-disubstituted indoles but are unstable for less substituted derivatives. Figure 1 indicates typical reactivity patterns toward oxygen.

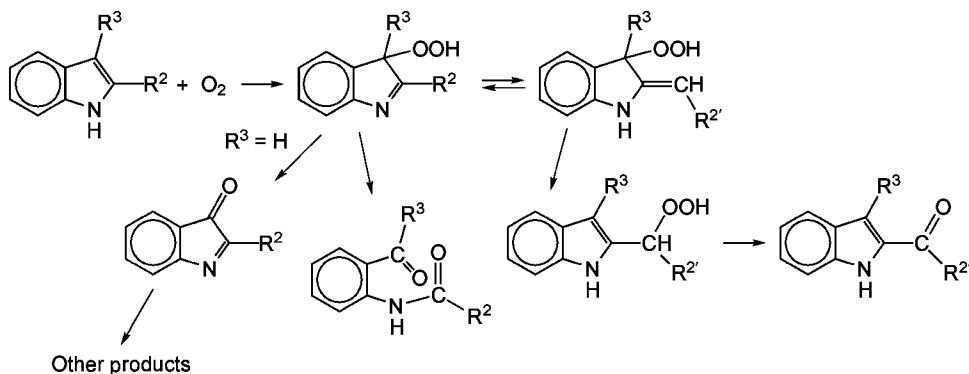
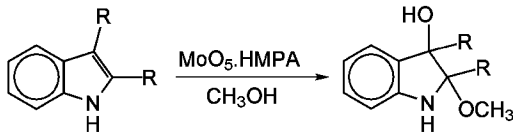


Fig. 1. Reactivity of 2,3-substituted indoles with oxygen.

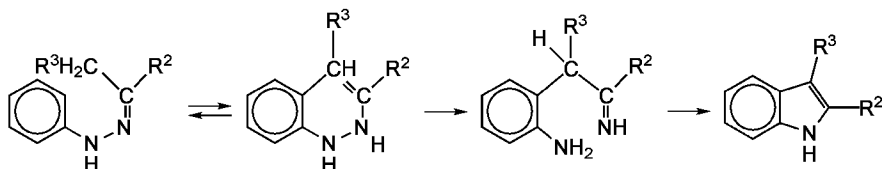
Mixtures of products are frequently observed. Oxidation by peroxycarboxylic acids usually give similar products (22). Several chemical oxidants give good yields of specific oxidation products. Dimethyl sulfoxide in aqueous acid gives oxindoles (23). In methanol, MoO₅·HMPA (hexamethylphosphoramide) gives 3-hydroxy-2-methoxyindolines (24).



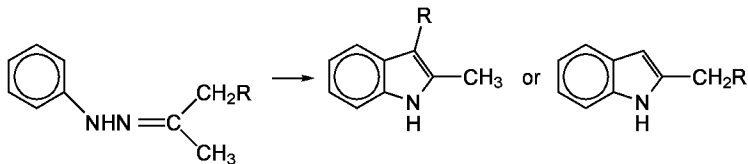
Syntheses

Although there are a wide variety of indole ring syntheses (25), most of the more useful examples fall within a small number of groups. Indole syntheses usually start with an aromatic compound, either monosubstituted or ortho-disubstituted. Those which begin with a monosubstituted starting material must at some point effect a substitution of the benzene ring.

The Fischer Indole Synthesis and Related Sigmatropic Syntheses. In the Fischer indole synthesis (26) an *N*-arylhydrazone is cyclized, usually under acidic conditions, to an indole. The key step is a [3,3]sigmatropic rearrangement of an enehydrazone tautomer of the hydrazone.

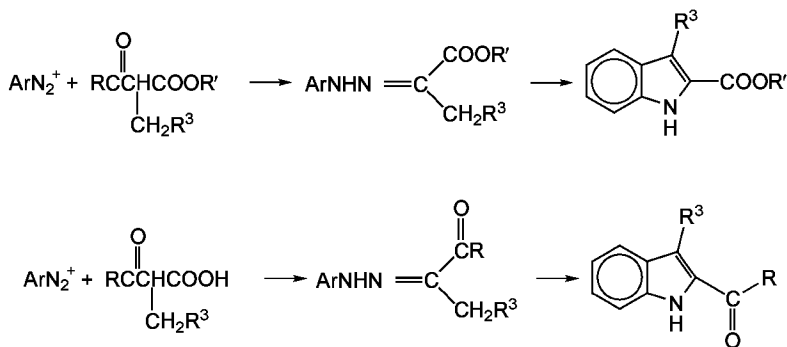


Thus to permit cyclization, there must be at least one hydrogen α to the C=N bond. If there is only one, the product will be a 3,3-disubstituted-3*H*-indole. However, if both substituents at the hydrazone carbon have one or more α -hydrogens, product mixtures can result. Generally, it is expected that the more branched substituent is more likely to be involved in cyclization, so typically phenylhydrazones derived from methyl alkyl ketones give 2-methylindoles. However, the selectivity is subject to the reaction conditions and with certain reagents the selectivity can be reversed to favor the 2-alkylindole.

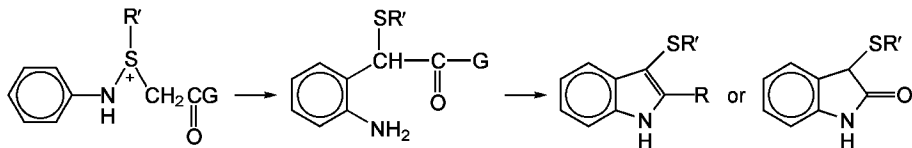


When the phenylhydrazone bears a meta-substituent, two isomeric indoles are possible; ortho-substituents also frequently introduce complications.

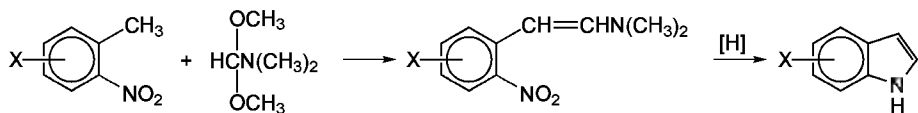
In addition to formation from a ketone, the hydrazones can be obtained from dicarbonyl compounds by a Japp-Klingemann reaction. This is especially useful for β -ketoesters and β -ketoacids, which undergo either decylation or decarboxylation.



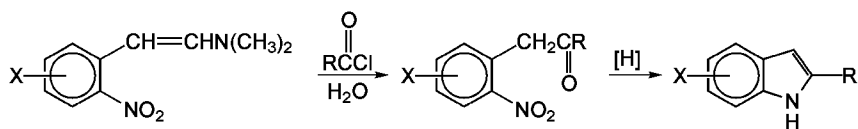
Another indole-oxindole synthesis achieves the critical ortho-substitution by Sommelet-Hauser rearrangement of an anilinosulfonium ion intermediate. Use of β -thioketones (G = R, an alkyl group) generates 2-substituted indoles, whereas β -thioesters (G = OR) lead to oxindoles. In each case, a 3-thio substituent must be removed by desulfurization.



Reductive Cyclizations. The Batcho-Leimgruber protocol involves condensation of an *o*-nitrotoluene with a dimethylformamide acetal to form a β -(*o*-nitrophenyl)enamine (27). A reducing agent then affects the reductive cyclization to an indole.

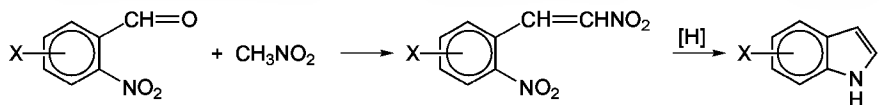


There have been a number of refinements to the procedure, both in the enamine formation and in the reduction. Furthermore, the procedure can be adapted to 2-substituted indoles by introducing an acyl substituent on the enamine intermediate.

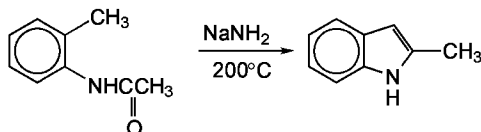


In general, any *o*-nitrobenzyl ketone or *o*-aminobenzyl ketone can be converted to a 2-substituted indole. There are a variety of specific examples of such syntheses, although there are not any truly general means of generating these kinds of starting materials.

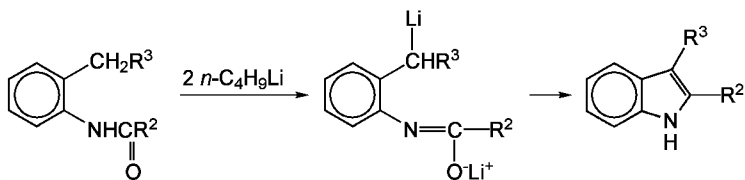
o-Nitrobenzaldehydes condense with nitromethane to give *o*, β -dinitrostyrenes. A variety of reducing agents convert these to indoles.



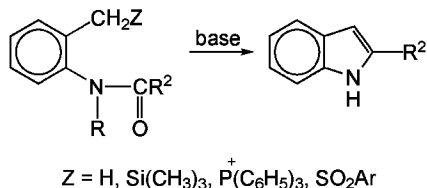
The Madelung Synthesis and Related Base-Catalyzed Condensations. The Madelung cyclization involves an intramolecular condensation of an *o*-alkylanilide. A classic example of the Madelung synthesis is the high temperature condensation of *o*-methylacetanilide [120-66-1] to 2-methylindole [95-20-5] by sodium amide.



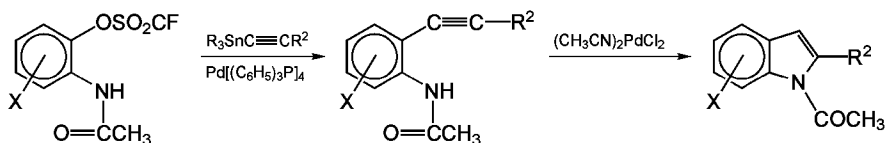
Cyclization can be achieved under much milder conditions by using *n*-butyllithium or lithium diisopropylamide to form a dilithio derivative of the anilide (28).



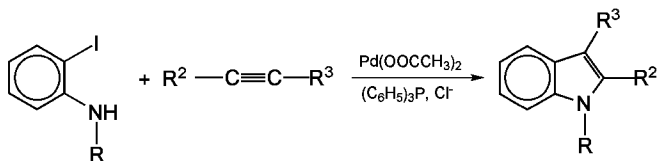
The same reactivity pattern is observed with *o*-methylanilides in which a carbanion-stabilizing substituent is attached to the methyl group. For Z = trimethylsilyl or triphenylphosphonio, elimination occurs with cyclization.



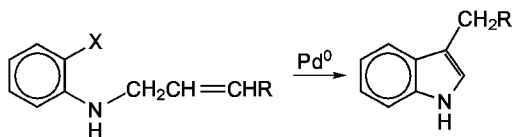
Transition-Metal Catalyzed Cyclizations. *o*-Halogenated anilines and anilides can serve as indole precursors in a group of reactions which are typically catalyzed by transition metals. Several catalysts have been developed which convert *o*-haloanilines or anilides to indoles by reaction with acetylenes. An early procedure involved coupling to a copper acetylide with *o*-iodoaniline. A more versatile procedure involves palladium catalysis of the reaction of an *o*-bromo- or *o*-trifluoromethylsulfonylanilide with a trialkylstannylalkyne. The reaction is conducted in two stages, first with a Pd(0) and then a Pd(II) catalyst (29).



o-Iodoaniline or *o*-iodoanilides can be cyclized to 2,3-disubstituted indoles by reaction with disubstituted alkynes in the presence of a Pd(II) catalyst (30). With unsymmetrical alkynes the bulkier group occupies the 2-position.

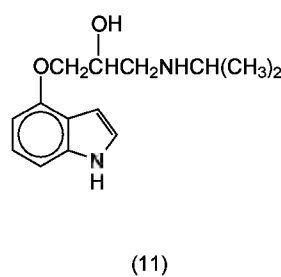
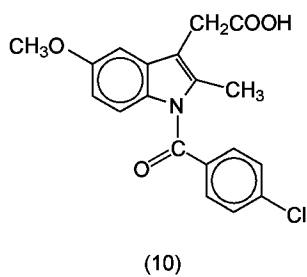


3-Substituted indoles can be prepared from *o*-bromo or *o*-iodoanilines by palladium-catalyzed cyclization of *N*-allyl derivatives (31).

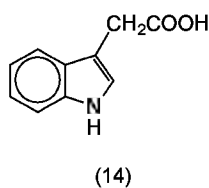
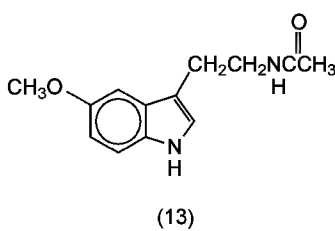
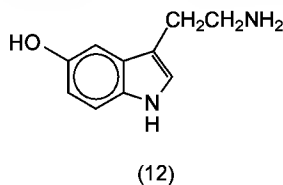


Biologically Active Indole Derivatives

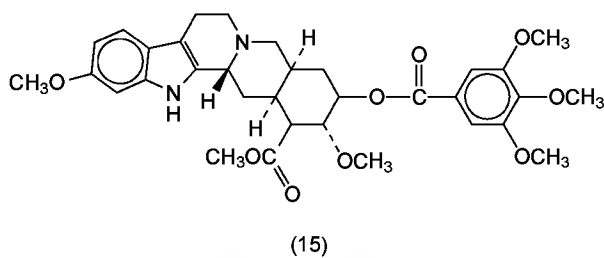
Synthetic Derivatives of Indoles as Pharmaceuticals. Thousands of indole derivatives have been prepared and evaluated as potential pharmaceuticals (32). Of those which have been put into use perhaps the most important are the nonsteroidal antiinflammatory agent indomethacin [53-86-1] (**10**) (33) and the β -adrenergic blocker pindolol [13823-86-9] (**11**) (34).



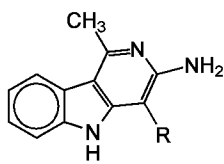
Naturally Occurring Compounds. Many derivatives of indole are found in plants and animals where they are derived from the amino acid tryptophan. Several of these have important biological function or activity. Serotonin [50-67-9] (12) functions as a neurotransmitter and vasoconstrictor (35). Melatonin [73-31-4] (13) production is controlled daily by the circadian cycle and its physiological level influences, and seasonal rhythms in humans and other species (36). Indole-3-acetic acid [87-51-4] (14) is a plant growth stimulant used in several horticultural applications (37).



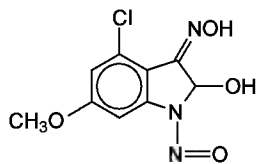
The largest single class of naturally occurring indoles are the plant alkaloids. These occur with a wide range of structural diversity and are typically derived from tryptophan and terpenoid structural units. Several of these compounds are pharmacologically significant. Reserpine [50-55-5] (15) acts as a tranquilizer and hypotensive agent. Although not widely used at present, it was one of the first drugs to be introduced for the treatment of mental illness (38). The dimeric vinca alkaloids vincristine [57-22-7] and vinblastine [865-21-4] are used in the treatment of Hodgkin's disease, leukemia, and other forms of cancer (39). Derivatives of the ergot alkaloid lysergic acid are used in the treatment of migraine (40) and the diethylamide is lysergic acid diethylamide (LSD) (see ALKALOIDS).



Toxic Indole Derivatives. There are several documented cases where indole derivatives, both natural and of synthetic origin, have been linked to pathological effects in humans. 3-Methylindole [83-34-1], which is produced by bacterial fermentation in cattle, can lead to pulmonary edema (41). The active metabolite is 3-methyleneindolenine, which forms adducts with proteins, primarily through sulfhydryl groups (42). The pyridoindoles (16) Trp-P-1 (R = CH₃) and Trp-P-2 (R = H) are genotoxic substances which originate from pyrolysis of tryptophan and have been identified in foods cooked at excessively high temperatures (43). 4-Chloro-6-methoxyindole, which can be extracted from fava beans, yields a potent mutagen on interaction with nitrite ion. This mutagen (17) has been associated with the etiology of stomach cancer in certain areas of Colombia (44). In the late 1980s a lethal pathological condition appeared which was associated with the consumption of L-tryptophan having a specific synthetic origin (45). The causative agent is believed to be a contaminant introduced by the specific manufacturing process. Several contaminants have been identified but the precise identity of the causative agent remains under investigation (46).



(16)



(17)

BIBLIOGRAPHY

"Indole" in *ECT* 1st ed., Vol. 7, pp. 839–847, by W. C. Sumpter, Western Kentucky State College, and F. M. Miller, University of Maryland; in *ECT* 2nd ed., Vol. 11, pp. 585–595, by W. C. Sumpter, Western Kentucky State College, and F. M. Miller, University of Maryland; in *ECT* 3rd ed., Vol. 13, pp. 213–222, by D. W. Bannister, Toms River Chemical Corp.

1. M. Balon and co-workers, *Tetrahedron* **45**, 7501 (1989).
2. A. W. Hanson, *Acta Cryst.* **17**, 359 (1964).
3. K. Gregory and co-workers, *Organometallics* **9**, 1485 (1990).
4. P. Roychowdhury and B. S. Basak, *Acta Crystallogr., Sect. B* **B31**, 1559 (1975).
5. P. Joseph-Nathan, R. E. del Rio, and M. S. Morales-Rios, *Heterocycles* **27**, 377 (1988).
6. M. S. Morales-Rios, J. Espineira, and P. Joseph-Nathan, *Magn. Reson. Chem.* **25**, 377 (1987).
7. R. Liu and Y. Zhang, *Xiandai Huagong* **9**, 18 (1989); T. Y. Gogoleva and S. N. Olshanskaya, *Koks Khim.*, 41 (1990).
8. Jpn. Kokai Tokkyo Koho 62 22, 758 (Jan. 30, 1987), T. Ueno and co-workers (to Mitsui Toatsu Chemicals, Inc.).
9. Jpn. Kokai Tokkyo Koho 59 172,467 (Sept. 29, 1984), (to Mitsui Toatsu Chemicals, Inc.).
10. Jpn. Kokai Tokkyo Koho 01 199,943 (Aug. 11, 1989), M. Tsuzaki and M. Sato (to Idemitsu Kosan Co., Inc.); **Brit. Pat. Appl. 2,232,152** (Dec. 5, 1990), E. Gerecs an co-workers (to Reanal Finomvegyszergyar).
11. Jpn. Kokai Tokkyo Koho 63 196,562 (Aug. 15, 1988), H. Sato and M. Tsuzuki (to Idemitsu Kosan Co., Inc.).
12. J. Catalan and co-workers, *J. Am. Chem. Soc.* **110**, 4105 (1988); J. Catalan and M. Yanez, *J. Am. Chem. Soc.* **106**, 421 (1984).
13. M. R. Brennan and co-workers, *Heterocycles* **24**, 2879 (1986); P. G. Gassman, G. A. Campbell, and G. Mehta, *Tetrahedron* **28**, 2749 (1972); M. De Rosa, L. Carbognani, and A. Febres, *J. Org. Chem.* **46**, 2054 (1981).
14. M. E. Flaugh and co-workers, *J. Med. Chem.* **22**, 63 (1979); D. Ranganathan and co-workers, *J. Org. Chem.* **45**, 1185 (1980).
15. R. A. Heacock and S. Kasperek, *Adv. Heterocycl. Chem.* **10**, 61 (1969).
16. J. Bergman and L. Venemalm, *Tetrahedron* **46**, 6061 (1990).
17. L. S. Hegedus, *Angew. Chem., Int. Ed. Engl.* **27**, 1113 (1988); T. Sakamoto, Y. Kondo, and H. Yamanaka, *Heterocycles* **27**, 2225 (1988).
18. M. A. Siddiqui and V. Sniekus, *Tetrahedron Lett.* **31**, 1523 (1990).
19. R. J. Sundberg and H. F. Russell, *J. Org. Chem.* **38**, 3324 (1973); M. G. Saulnier and G. W. Gribble, *J. Org. Chem.* **47**, 757 (1982).
20. I. Hasan and co-workers, *J. Org. Chem.* **46**, 157 (1981).
21. A. R. Katritzky and K. Akutagama, *Tetrahedron Lett.* **26**, 5935 (1985).
22. E. Braudeau, S. David, and J.-C. Fischer, *Tetrahedron* **30**, 1445 (1974).
23. K. Szabo-Pusztay and L. Szabo, *Synthesis*, 276 (1979).
24. C.-S. Chien and co-workers, *Chem. Pharm. Bull.* **34**, 1493 (1986).
25. U. Pindur and R. Adam, *J. Heterocyclic Chem.* **25**, 1 (1988); M. A. Yurovskaya, *Chemistry of Heterocyclic Compounds (Engl. Transl.)* **23**, 919 (1987).
26. B. Robinson, *The Fischer Indole Synthesis*, John Wiley & Sons, Inc., New York, 1982.
27. R. O. Clark and D. B. Repke, *Heterocycles* **22**, 195 (1984).
28. W. J. Houlihan, V. A. Parrino, and V. Uke, *J. Org. Chem.* **46**, 4511 (1981).
29. D. E. Rudisill and J. K. Stille, *J. Org. Chem.* **54**, 5856 (1989).
30. R. C. Larock and E. K. Yum, *J. Am. Chem. Soc.* **113**, 6689 (1991).
31. R. Odle and co-workers, *J. Org. Chem.* **45**, 2709 (1980).
32. K. C. Joshi and P. Chand, *Die Pharmazie* **37**, 1 (1982); J. K. Sugden and T. O. Yoloze, *Pharm. Acta Helv.* **53**, 65 (1978).
33. T. Y. Shen and C. A. Winter, *Adv. Drug Res.* **12**, 89 (1977).
34. W. H. Frishman, *N. Engl. J. Med.* **308**, 940 (1983).
35. N. N. Osborne, ed., *Biology of Setonergic Transmission*, John Wiley & Sons, Inc., Chichester, U.K., 1982; P. M. Van Loutte, R. A. Cohen, and J. M. Van Neuten, *J. Cardiovasc. Pharmacol.* **6**, (Suppl. 2), 421 (1984).
36. S. M. Armstrong, *Experientia* **45**, 932 (1989); T. J. Bartmess and D. B. Goldman, *Experientia* **45**, 939 (1989).
37. E. A. Schneider and F. Wrightman, *Phytobom. Related Compd; Compr. Treatise* **1**, 29 (1978).
38. E. Schlittler, *The Alkaloids* **8**, 231 (1965).
39. A. Brossi and M. Suffness, *The Alkaloids* **37**, 1 (1990).
40. I. Ninomiya and T. Kiguki, *The Alkaloids* **38**, 1 (1990).
41. J. R. Carlson and co-workers, *J. Toxicol., Toxin Rev.* **5**, 217 (1986).
42. W. Ruangyutikarn, G. L. Skiles, and G. S. Yost, *Chem. Res. Toxicol.* **5**, 713 (1992).
43. Y. Hashimoto, K. Shudo, and T. Okamoto, *Acc. Chem. Res.* **17**, 403 (1984).
44. D. Yang and co-workers, *Carcinogenesis* **5**, 1219 (1984).
45. E. A. Belongia and co-workers, *N. Engl. J. Med.* **323**, 357 (1990).
46. M. J. Smith and co-workers, *Tetrahedron Lett.* **32**, 991 (1991); Y. Goda and co-workers, *Chem. Pharm. Bull.* **40**, 2236 (1992).

General References

C. W. Bird and G. W. H. Cheeseman, eds., *Comprehensive Heterocyclic Chemistry*, Vol. 4, Pergamon Press, Oxford, 1984, Chapt. 3.04, 3.05, and 3.06.

W. J. Houlihan, ed., *The Chemistry of Heterocyclic Compounds*, Vol. 25, Parts 1, 2, and 3, Wiley-Interscience, New York, 1972.
R. J. Sundberg, *The Chemistry of Indoles*, Academic Press, Inc., New York, 1970.

Richard J. Sundberg
University of Virginia

INDOPHENOL.

See SULFUR DYES.

INDULINES.

See AZINE DYES.

INDOPHENOL.

See SULFUR DYES

INDULINES.

See AZINE DYES

INDUSTRIAL ANTIMICROBIAL AGENTS

Industrial antimicrobial agents are chemicals used to prevent the adverse consequences of microbiological activity in processes and products. Some are unique to this segment and others are drawn from the antimicrobial agents used in medicine, agriculture, and sanitary applications. Industrial antimicrobials are selected where process or strictly physical conditions, such as irradiation or heat, are impractical or ineffective in controlling microbiological activity.

Microorganisms are ubiquitous, thus microbial contamination is the rule; the total absence of microbes, ie, sterility, is the exception. Many microorganisms might be considered mainstream, growing under typical ambient conditions, but there are almost always strains that are capable of surviving and multiplying under the extremes of pH, salinity, pressure, and temperature.

Chemical suppliers include basic manufacturers of active ingredients, formulators, and distribution or service industries. The relative importance of each depends greatly upon the industry being supplied. In many instances, the vendor may supply a number of performance chemicals (eg, corrosion control agents or stabilizers) in addition to the antimicrobial agent.

Microbiology

Industrial antimicrobial agents are targeted against bacteria, mold, and fungi. Bacteria are simple, single-celled organisms that are rarely over two to three micrometers in size. The identification of a microorganism is based on its size, shape, biochemical activity, and growth characteristics. More recently, the fields of serology and DNA homology have contributed to rapid and simplified bacterial identification.

Fungi are somewhat more complex than bacteria and more similar in physiology to the higher organisms. This similarity, unfortunately, makes discovery of biocides that specifically block fungal growth, with no effects harmful to humans, much more difficult. Fungal cells typically differentiate so that not all of the cells making up the organism are the same, which is not true of bacteria. Fungi readily grow into a size easily visible to the unaided eye.

Yeasts and algae are also microorganisms that fall within the category of targets of industrial antimicrobial agents. Algae, of course, are photosynthetic, that is, they convert light into chemical energy. Yeasts are similar to fungi in physiology, but are unicellular like bacteria.

All organisms require a source of energy, nitrogen for nucleic acid and amino acid synthesis, and essential nutrients such as phosphorus, sulfur, magnesium, and water. Microorganisms display a tremendous range in their requirements for nutrients. Fastidious organisms require specific forms of certain nutrients, such as amino acids or vitamins. Organisms causing problems in industrial situations are frequently at the opposite end of the spectrum, being extremely self-sufficient in synthesizing required biochemicals from the most basic molecules. The type and availability of nutrient may be the most significant determinant in the microbiological ecology of the products and processes which require the use of an industrial antimicrobial agent.

Viruses are obligate intracellular parasites. They only exhibit activity by infecting other living organisms, thus they are not a practical concern in industrial microbiological fields. The exception is where viral contamination of the product or process represents a threat of transmission of disease. Microscopic insects and protozoans are also not addressed in this article (see INSECT CONTROL TECHNOLOGY).

There are about 30,000 species of bacteria (1), which, by historical convention, are placed into gram-positive or gram-negative groups (after Hans Gram, a turn-of-the-century Danish researcher), based on the bacterial affinity for staining with a dye (2). Both gram-positive and gram-negative organisms are problematic in a number of industrial situations. They display some distinctiveness in their individual responses to antimicrobial agents.

Antimicrobial Activity

Microbiologists have developed the following hierarchy to categorize the expected level of performance of antimicrobial treatments:

Category	Result of treatment
sterilant	completely kills all life forms
sporicidal	kills spores
disinfectant	kills all infectious bacteria
cidal	kills all organisms of type (eg, tuberculocidal)
sanitizer	reduces number of organisms to safe level
antiseptic	prevents infection
static	prevents growth of organism (eg, fungistatic)

Some chemicals function by killing microorganisms (3,4), whereas others merely prevent organisms from growing or reproducing. Frequently biocide chemicals with a high degree of cidal activity are reactive. Clearly, if the reactivity is not specifically targeted toward biochemicals essential to the microorganisms in the system, the biocide may be consumed in the application and rendered ineffective. Biocide chemicals limited to a static level are typically less reactive and exhibit a longer span of functionality. Thus, cidal activity is not necessarily preferred to static activity. The selection of biocide chemistry is a balance between reactivity and stability, which must be attuned to the requirements of the industrial application. A more extensive survey of the nomenclature associated with levels of microbiological control has been provided (5) (see DISINFECTANTS AND ANTISEPTICS).

Measures of Activity. The potency of an antimicrobial to kill or render the organisms inactive is measured by the minimum inhibitory concentration (MIC). This property is obtained in a laboratory test in which the test compound is added at stepwise concentrations to microbiological media, followed by inoculation with a pure culture of microorganism. In the absence of growth, the media are unchanged. If growth occurs, the media appear turbid due to the high population of the bacteria. The lowest concentration inhibiting growth is the MIC for that organism (the lower the number, the more potent the antimicrobial). MICs are widely available, but minor differences in protocol result in some reported differences. The MIC is of little value to the users of antimicrobials because it is not a true predictor of performance in any particular application, and more application-specific tests have been developed.

Applications

Practically, the broad general area of industrial antimicrobial concerns is divided into process and preservative disciplines. Figure 1 shows the key application areas of process and companion-preservative uses, and Table 1 indicates the various market segments.

Process uses	Preservative uses
Swimming pool sanitizers	
Cooling water	
Metalworking	
Pigment slurries	
Petroleum recovery	Jet fuel
Pulp	Paper
Paint (in-can)	Paint film
Latex	Caulks
Adhesive	Adhesives
Hide processing	Leather
Sapstain	Wood
Laundry sanitizers	Textile
	Plastics
	Cosmetics

Fig. 1. Applications of industrial antimicrobial agents.

Table 1. Value of U.S. Antimicrobial Industrial Agent Markets

Market segment	Value, 10 \$
wood sapstain	200
swimming pool sanitizers	200
paints coatings	50-100
cooling water	20-50
metalworking	20-50
pulp paper	20-50
plastics	20-50
adhesives	2-20
petroleum fuel	2-20
slurries	2-20
latex caulk	2-20
laundry textile	2-20
hides leather	2-20

It is common to consider the preservative use a service-life challenge which may require several years for successful performance, whereas the process uses are of limited duration. On the other hand, some processes are ongoing operations which require regular treatment with antimicrobials. The commonality running throughout the various processes is the presence of significant amounts of water. Both bacteria and fungi can flourish in aqueous systems; fungi are most notably the problem when liquid moisture is less available, as might be encountered during the service life of the products listed as preservatives in Figure 1.

Efficiency testing needs to be conducted with actual conditions simulated as closely as possible (3,6,7). Table 2 presents some of the standardized protocols available to microbiology labs for confirmation of usefulness (8).

Table 2. Common Test Methods for Evaluating Materials^a

Material	Method	Organism	Result
adhesives	ASTM D1174	mixed bacterial culture	bacterial growth, viscosity change
	ASTM D4300	mixed fungi	growth on film
	ASTM D1286	mixed fungal culture	fungal growth, viscosity change
metalworking fluid	ASTM E686	fungal and bacterial mixed culture	organism growth, physical change
	ASTM D3946	spoiled material	organism growth, physical change
	ASTM D2574	bacterial culture	bacterial growth

paper	ASTM D3273	mixed fungal culture	surface growth
	UL181	fungal culture	surface growth
plastic	ASTM D2020	mixed fungal culture	surface growth
	TAPPI T487, M54	mixed fungal culture	surface growth
	ASTM G21	mixed fungal culture	surface growth
	ASTM G22	<i>Pseudomonas aeruginosa</i>	visible growth
textile	ASTM G29	freshwater algae	algal attachment
	ASTM D3083	soil burial	staining, weight loss, stiffness
	AATCC Method 90	bacterial cultures	zone of inhibition
	AATCC Method 100	bacterial cultures	percent reduction
wood	ASTM D2017	fungal enriched soil burial	weight loss
	ASTM D1413	fungal enriched soil burial	weight loss, decay

^a Ref. 8

Regulation of Antimicrobial Agents

The key trend affecting this industry is the regulation of products and their uses by the U.S. Environmental Protection Agency (EPA). This regulation has curtailed the introduction of new ingredients, selectively limited the markets for a number of products, and eliminated products by suspension or cancellation. The data in Figure 2 show the number of approvals by the EPA made for new industrial antimicrobials since the early 1970s (9). During this same period, the total number of new active ingredients (eg, insecticides, agricultural fungicides, and herbicides) has averaged between 12 and 13 approvals per year. The hallmark of approval is a registration.

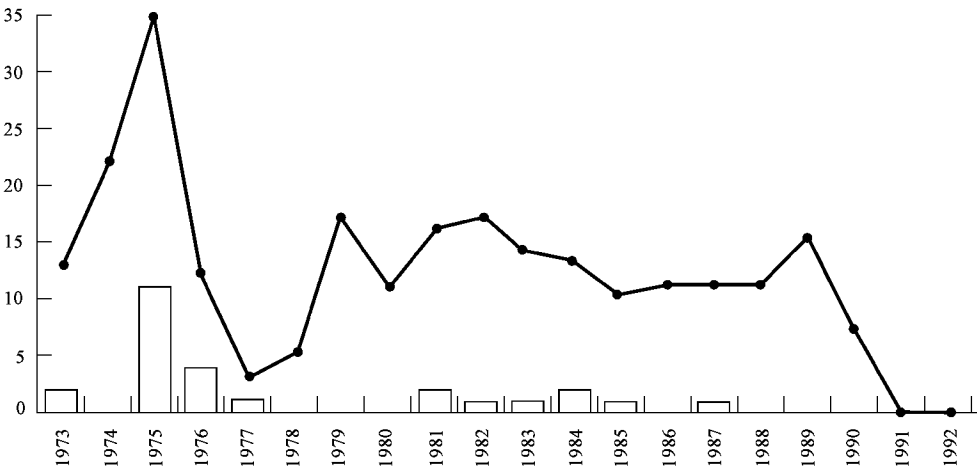


Fig. 2. EPA registrations issued for new active ingredients: (□), new bactericides and slimicides; (●), all pesticides.

Several antimicrobials have been banned or severely restricted by the EPA based on documented or suspected toxicity or environmental problems. Others have been discontinued in the face of testing costs required by the EPA reregistration program mandated by the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) of 1988 (10). Some of the significant products that have become obsolete are 2,4,5-trichlorophenol[95-95-4], sodium 2,4,5-trichlorophenate[136-32-3], hexachlorophene[70-30-4], 2,2'-methylenebis(4-chlorophenol)[27-23-4], phenylmercuric acetate[62-38-4], and 3,4,5-tribromosalicylanilide[38848-67-8].

Contraction in the number of EPA-allowed biocides has heightened efforts to develop naturally derived preservatives and microorganisms capable of countering microbial degradation. Neem oil (*Azadirachta indica* seed extract) has been featured as an exceptional natural candidate for the preservation of cosmetic products. Naturally derived chemicals with antimicrobial properties have been used since antiquity as preservatives. However, displacement of successful synthetic products by natural products in preservatives of any category remains to be witnessed.

Toxicology. Industrial antimicrobial agents are regulated in the United States as pesticides under FIFRA. Thus the industrial antimicrobial regulation is strongly influenced by agricultural pesticide considerations. The hallmark of the EPA allowance is registration, which is the culmination of a process that begins with the submission of toxicology, chemistry, and environmental data showing that the product can be used without "adverse effects on man and the environment." An approval to sell a product in the United States reflects a review of substantial amounts of toxicology data and a conclusion that the product can be safely used for its intended purpose.

The most often quoted toxicology data is the acute oral LD₅₀. This number is a crude measurement of the potential hazard to humans (and other mammals) from ingestion. Ingestion by individuals working in industrial commercial settings is limited as a result of training programs, warnings, and industrial hygiene; therefore the LD₅₀ alone bears little direct relevance to the safe use of a product. Much more instructive is the documentation of quantifiable hazards associated with repeated exposure scenarios that are typical of the conditions of use (11).

The U.S. EPA regulates both the active ingredient (technical or formulating grade) and the formulated (end use) products. Data requirements for the active ingredient are much more substantial than for the formulated product. Because of the importance of toxicology studies to the development of antimicrobials, the minimum toxicology data requirements and cost estimates for studies conducted at contract laboratories are noted in Table 3 (12). This data is required for products with minimal opportunity for human exposure, Tier I (according to EPA convention). Analysis of the data obtained in the Tier I series allows a rational comparison of favorable or unfavorable toxicology of a candidate antimicrobial agent.

Table 3. Estimated Cost of Toxicology Study^a

Study	Cost for active ingredient, \$	Cost for formulation, \$
<i>Acute series</i>		
acute dermal LD ₅₀	7,000	7,000
acute inhalation LC ₅₀	22,000	22,000
eye irritation	1,250	1,250
skin irritation	1,250	1,250
sensitization	10,000	10,000
<i>subtotal</i>	45,000	45,000

	<i>Tier I^b series</i>	
90-day subchronic	75,000	
teratology	125,000	
mutagenicity	30,000	
<i>subtotal</i>	230,000	45,000
<i>Total</i>	275,000	45,000

^a Ref. 12.

^b EPA-defined as minimal opportunity for human exposure.

The cost of developing a new active ingredient, however, is much more costly than the basic costs involved in toxicology studies, as shown in Table 3, and is likely to be \$35 × 10⁶ (13).

Reregistration. In addition to its authority to control the availability of new active ingredients, formulations, and the way in which the formulas are used, EPA is reapproving the use of older products through a process known as reregistration. In order to achieve some administrative economy, EPA has grouped some of the biocides with chemical similarity into a generic cluster known in reregistration as a case. Additionally, there are a number of active ingredient families which are being handled in a generic case. For example, the case of tributyltin- (TBT) containing compounds includes bis(tributyltin) oxide[56-35-9], bis(tributyltin) salicylate[22330-14-9], tributyltin benzoate[4342-36-3], tributyltin fluoride[1983-10-4], tributyltin maleate[14275-57-1], and tributyltin methacrylate[2155-70-6].

Because manufacturers are unwilling to generate data to substantiate the acceptability of continued use, the other tin compounds of the case have become obsolete, including bis(tributyltin) adipate[7437-35-6], bis(tributyltin) sulfosalicylate[4419-22-1], tributyltin acetate[56-36-0], tributyltin acrylate[13331-52-7], tributyltin chloride[1461-22-9], tributyltin linoleate[24124-25-2], tributyltin monopropylene glycol maleate[53466-85-6], and tributyltin neodecanoate[28801-69-6]. Because of the substantial number of tin compounds, this is not a typical example, but it does illustrate the point that there has been a large decline in the availability of niche products.

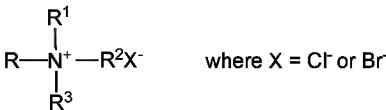
Manufacturers and Suppliers

Manufacturers and suppliers of industrial antimicrobial active ingredients and formulations include the following companies:

ANGUS Chemical Co.Northbrook, Ill.	Miles, Inc.Pittsburgh, Pa.
BASF Parsippany, N.J.	Morton International, Inc.Danvers, Mass.
Calgon Corp.Moon Run, Pa.	Petrolite Corp.St. Louis, Mo.
Dearborn Division,W. R. GraceLake Zurich, Ill.	Stepan Co.Northfield, Ill.
Huls AmericaPiscataway, N.J.	Union Carbide ChemicalsDanbury, Conn.
Lonza Inc.Fairlawn, N.J.	Baker Performance ChemicalsHouston, Tex.
Mooney Chemical, Inc.Cleveland, Ohio	Buckman LaboratoriesMemphis, Tenn.
Osmose Wood Perserving Co.Buffalo, N.Y.	Dow Chemical USAMidland, Mich.
Rohm and Haas Co.Philadelphia, Pa.	Hickson Corp.Conley, Ga.
Troy Chemical Corp.Newark, N.J.	Lehn & Fink Products Co.Montvale, N.J.
Zeneca (formerly ICI)Wilmington, Del.	Monsanto Co.St. Louis, Mo.
Atochem North AmericaPhiladelphia, Pa.	Olin Corp.Cheshire, Conn.
Betz LaboratoriesTrevose, Pa.	Rhône-Poulenc RorerPrinceton, N.J.
CSCharlotte, N.C.	SherexDublin, Ohio
Great Lakes ChemicalsWest Lafayette, Ind.	Vinings Industries, Inc.Atlanta, Ga.
ISK BiotechMentor, Ohio	

Classes of Antimicrobial Compounds

Quaternary Ammonium Compounds. These compounds (quats) have the following general formula:



The quaternary ammonium compounds (qv) are manufactured by the reaction of an alkyl halide with a tertiary amine. The alkyl halide may be short-chain, long-chain, or benzyl. Selection of a long-chain alkyl group yields structures with variable composition and greatly adds to the chemical complexity inherent in this group. Investigation of structure-activity relationships has led to claims for superior efficacy or compatibility, most notably with anionic surfactants in disinfectant-detergent cleaner systems, of closely related compounds in the family.

The quats are an extremely important group in medical and sanitary applications, with comparatively limited industrial applications. Activity against bacteria of public health importance is absolutely required, with a lesser demand being made for antifungal activity.

The key markets for quaternaries are as swimming pool algacides and in cooling water applications (see WATER, TREATMENT OF SWIMMING POOLS, SPAS, AND HOT TUBS), which further explains their importance as process biocides rather than preservatives. Some uses in latex films and plastics have been claimed (14,15). Primary quaternary ammonium industrial antimicrobial agents and their producers are presented in Table 4.

Table 4. Quaternary Ammonium Industrial Antimicrobial Agents

Chemical name	CAS Registry Number	Structure	R	Trade name	Producer
benzalkonium chlorides	[68391-01-5]	$\begin{array}{c} \text{CH}_3 \\ \\ \text{C}_6\text{H}_5\text{CH}_2-\text{N}^+-\text{(CH}_2\text{)}_{12-18}\text{Cl}^- \\ \\ \text{CH}_3 \end{array}$	(C ₁₂₋₁₈ H ₂)	Variquat	Huntington
	[68424-85-1]		(C ₁₂₋₁₈ H ₂)	BTC	Lonza

	[139-08-2]		(C ₁₄ H ₂)		Sherex Stephan
dialkyldimethyl-ammonium chlorides	[5538-94-3]		C ₈	Bardac LF	Lonza Stephan
	[7173-51-5]		C ₁₀	Bardac 2250	
cetyltrimethyl-ammonium bromide	[57-09-0]		C ₁	Bardac 22 Bromat	Lonza Hexcel
cetylpyridinium chloride ^a	[123-03-5]		C ₁		Hexcel
3-(trimethoxysilyl)-propyldimethyl-octadecyl-ammonium chloride	[27668-52-6]		C ₁₈	Q-5700 Durotex QST	Dow-Corning Morton

^a Cetyl = hexadecyl.

The mechanism of action of quats has been widely studied. It is generally agreed that their interaction with the bacterial cell membrane is the primary event resulting in antimicrobial activity (16,17).

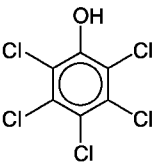
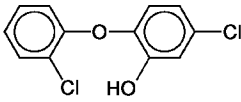
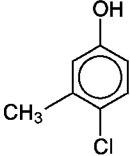
Regulatory pressures which have limited the introduction of new chemistries have led to a substantial trend to obtain innovation through combination. One apparently successful introduction is the combination of a quat with a polyphase for sapstain prevention on freshly sawn lumber (18).

Quats are usually moderately soluble in water, but this varies widely owing to the range of groups bonded to the nitrogen. They are fundamentally nonreactive but act as surface-active cations. Compatibility with anionic detergents and activity in the presence of hard water have been claimed for some quats (19).

Phenolics. Phenol (qv) and the chlorinated phenolics formerly comprised the largest class of industrial antimicrobials (see CHLOROPHENOLS). Table 5 shows the remaining phenolics of importance. Use of pentachlorophenol has been severely restricted; only one manufacturer supplies product for the wood preservation market.

Table 5. Phenolic Industrial Antimicrobials

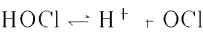
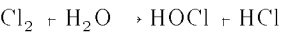
Chemical name	CAS Registry Number	Structure	Trade name	Producer	Applications
2-benzyl-4-chlorophenol	[120-32-1]		Santophen I	Monsanto	disinfectants
<i>o</i> -phenylphenol	[90-43-7]		Dowicide 1	Dow	metalworking fluids, leather, paint
sodium <i>o</i> -phenylphenate	[132-27-4]		Dowicide A	Dow	metalworking fluids, adhesives, paint, textiles
pentachlorophenol	[87-86-5]		Glazd	Vulcan	wood

					
2-(2',4'-dichlorophenoxy)-5-chlorophenol	[3380-34-5]		Irgasan	CIBA-GEIGY	textiles
4-chloro-3-methylphenol	[59-50-7]		PCMC	Hbbs	adhesives, latex

Sodium orthophenyl phenate remains an important ingredient in the treatment of cooling waters. Both orthophenyl and its sodium salt have a wide spectrum of preservative use, including caulks, construction products, and leather processing. Similarly, *para*-chloro-*meta*-xylenol is used to preserve a number of water-based goods, including inks (qv), emulsions, and shoe polish, in addition to providing mold resistance to leather.

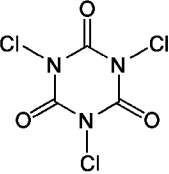
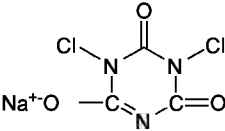
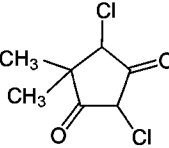
The most common use of 2-(2',4'-dichlorophenoxy)-5-chlorophenol (2,4,4'-trichloro 2'-phenoxyphenol) is in the personal care products market, where it is commonly known as triclosan and is the active antibacterial in underarm deodorants. It has also found some acceptance as an antibacterial component of plastic mattress covers.

Chlorine and Bromine Oxidizing Compounds. The organo chlorine compounds shown in Table 6 share chemistry with inorganic compounds, such as chlorine[7782-50-5] and sodium hypochlorite[7681-52-9]. The fundamental action of chlorine compounds involves hydrolysis to hypochlorous acid (see CHLORINE OXYGEN ACIDS AND SALTS).



The dissociation of hypochlorous acid is an equilibrium reaction and pH-dependent. The microbiological activity of chlorine and its derivatives is strongly influenced by the pH of the solution, exhibiting much greater activity at pH less than 7, leading to the conclusion that the active microbiocide is the undissociated hypochlorous acid (20). Principal halogen-containing industrial antimicrobial agents and their manufacturers and applications are listed in Table 6. Iodine compounds include *p*-tolylidiodomethyl sulfone[20018-09-1] (Amical 48), a paint mildewcide produced by Angus, and 3-iodo-2-propynylbutyl carbamate [55406-55-6] (Polyphase), used on wood and in adhesives, produced by Troy.

Table 6. Halogen-Containing Industrial Antimicrobial Agents

Chemical name	CAS Registry Number	Structure	Trade name	Producer	Applications
<i>Chlorine compounds</i>					
trichloroisocyanurate	[87-90-1]		ACL-85 CDB-90 Pace	Monsanto FMC Olin	swimming pool sanitizer
sodium dichloroisocyanurate	[2893-78-9] ^a		ACL-60 ACL-56 Clearon	Monsanto Monsanto FMC	swimming pool sanitizer, disinfectant
potassium dichloroisocyanurate	[2244-21-5]	Na ⁺ O ⁻	ACL-59	Monsanto	swimming pool sanitizer, disinfectant
monotrichloroisocyanurate: potassium dichloro-isocyanurate, 1:4 dichlorodimethylhydantoin	[34651-95-1] [118-52-5]		ACL-66 Halane	Monsanto BASF Wyandotte	swimming pool sanitizer, disinfectant
<i>Bromine compounds</i>					
bromochlorodimethylhydantoin	[126-06-7]		DiHalo sticks	Great Lakes	swimming pool disinfectant
2,2'-dibromo-3-nitropropionamide e	[10222-01-2]	N≡CCBr ₂ CONH ₂	DBNPA	Dow	cooling water slimicide, secondary oil recovery metalworking

bis(1,4-bromoacetoxy)-2-butene	[20679-58-7]	BrCH₂COOCH₂C H BrCH₂COOCH₂C H	Slimacide V-10	Vineland	fluids paper-mill slimicide
1,2-dibromo-2,4-dicyanobutane	[35691-65-7]	BrCH ₂ CBr(CN)CH ₂ CH ₂ CN	Tektamer 38	Calgon	paint, adhesive, latex emulsion, joint compound
2-bromo-2-nitropropane-1,3-diol	[52-51-7]	HOCH ₂ CBr(NO ₂)CH ₂ OH	Bronopol	Angus	cosmetic
benzyl bromoacetate	[5437-45-6]	C H ₅ CH ₂ —OOCCH ₂ Br	Merbac 35	Calgon	preservative paint, latexes

^a The dihydrate has the CAS Registry Number [57580-86-0]

The organo chlorine compounds are more expensive than inorganic chlorine compounds, but offer improved stability against photolytic breakdown in swimming pools (21). Swimming pool sanitation is generally accomplished with 1–3 ppm free chlorine residual (see CHLORAMINES AND BROMAMINES; WATER, TREATMENT OF SWIMMING POOLS, SPAS, AND HOT TUBS).

Inorganic chlorine and organic chlorine compounds are also used in cooling water treatment. The cost of organic chlorine is offset by a greater range of pH tolerance (22). The efficacy of chlorine compounds is due to strong oxidizing capacity. However, because of the nonspecific nature of the reactivity, a significant portion of the added chemical can be lost to reaction with organic matter (eg, partially decayed chemicals of biological origin) contaminating the system. Thus it is necessary to satisfy the chlorine demand of the system (23).

In cooling water applications, great importance is placed on activity against *Legionella pneumophila*, the causative agent of Legionnaires' disease. Bromochloro dimethylhydantoin has been shown to rapidly hydrolyze in water with the formation of hypobromous acid (21). The p*K*_a of hypobromous acid is 8.8, whereas the p*K*_a of hypochlorous acid is 7.4. Because the undissociated hypohalous acid is the active biocide, the hypobromous-generating chemical is more active in alkaline systems.

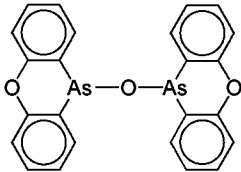
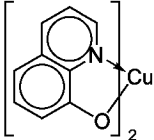
Chlorine dioxide has substantial reactivity, which precludes its shipment in bulk. New technology that allows on-site generation of ClO₂ from sodium chlorate [7775-09-9] rather than from chlorine is expected to result in its more frequent use in applications where capital investment and operators are warranted (24).

Sodium bromide is the most rapidly growing antimicrobial in water treatment applications (25). Chlorine dioxide [10049-04-4] has not been historically important, but may have a bright future because of its excellent antimicrobial activity without formation of halomethanes or chloramines (26).

Resistance to antimicrobial agents is of concern as it is well known that bacterial resistance to antibiotics can develop. Many bacteria already derive some nonspecific resistance to biocides through morphological features such as their cell wall. Bacterial populations present as part of a biofilm have achieved additional resistance owing to the more complex and thicker nature of the biofilm. A system contaminated with a biofilm population can require several orders of magnitude more chlorine to achieve control than unassociated bacteria of the same species. A second type of resistance is attributed to chemical deactivation of the biocide. This deactivation resistance to the strong oxidizing biocides probably will not occur (27).

Organometallics. Organometallics, listed in Table 7, were previously dominated by the mercurial compounds. For example, phenyl mercuric acetate [62-38-4] previously was a dominant mildewcide for paint (28), but was regulated out of use by EPA in 1990. Mercury compounds were preferred for their broad activity and effectiveness, both for in-can and paint-film protection, compared to competing products. The industry had known of the perilous status of mercurials prior to cancellation; however, the focus of work was on the concern for mercury entering the environment through manufacture and disposal of waste (29). The undoing of mercury was the result of the inability to contend with predictable misuse, eg, overformulating and use of exterior paint indoors, where indoor air contamination and consequent human intoxication by inhalation could occur.

Table 7. Metal-Containing Industrial Antimicrobial Agents

Chemical name	CAS Registry Number	Structure	Trade name	Producer	Applications
10,10'-oxybisphenoxarsine	[58-36-6]		Vinyzene Durotex	Morton	plastics, textiles
tributyltin oxide	[56-35-9]	(<i>n</i> -C ₂ H ₉) ₃ Sn—O—Sn(<i>n</i> -C ₄ H ₉) ₃	Fungitrol	HbIs	antifoulant, paint, plastics
tributyltin fluoride	[1983-10-4]	(<i>n</i> -C ₄ H ₉) ₃ SnF	Biomet 204	M&T	antifoulant paint
copper 8-quinolinolate	[10380-28-6]		Cunilate Nytek	Morton	textiles, wood
copper naphthenate	[1338-02-9]	Cu ₃ (AsO ₄) ₂ NaCr ₂ O ₇	Nuodex	Mooney	textiles, wood
chromated copper arsenate	[37337-13-6]			Kopcote	wood
ammoniacal copper arsenate	[32680-29-8]		All Weather Wood	Osmose	wood
cuprous oxide	[1317-39-1]			American-Chemet SCM-Glidden	antifoulant paint

The compound 10,10'-oxybisphenoxarsine (OBPA) is a unique organoarsenic antimicrobial. The powerful antimicrobial properties of organoarsenics had long been recognized, and there is plentiful treatment of the synthesis of organoarsenic derivatives in the literature. OBPA is used primarily in the plastics industry, where its high temperature stability and low vapor pressure make it the leading industrial antimicrobial in this market (30).

Use of organotin compounds in marine antifouling applications is governed by the Organotin Antifouling Paint Act of 1988, which limits use to ship hulls greater than 25 m in length and a maximum release rate of 2 $\mu\text{g cm}^2 \text{ d}$. Organotins are almost always used in combination with some other toxicant (31). Similarly, in water treatment, organotins are used as combination treatment (32).

Copper quinolinolate (oxine copper) is the chelate of divalent copper and 8-hydroxyquinoline and shares most of its market with copper naphthenate, which is a complex copper salt of mixed naphthenic acids. The principal uses are in wood treatments and some military textiles, where the green color is not objectionable. Copper naphthenate has an odor but is cheaper than oxine. Both copper naphthenate and zinc naphthenate have performed well in environment tests, with exposure to soil above-ground, as well as concrete (33).

A variety of copper soaps and complexes, eg, copper oleate [1120-44-1] and copper ethylenediamine tetraacetate [54430-03-1] (EDTA), have also been promoted, but the existence of these minor products is in doubt as a result of the EPA reregistration program mentioned previously. There is an exceptionally rich heritage of organometallic chemistry in the literature which does not have present commercial usefulness but is historically significant (34). Barium metaborate [13701-59-2], $\text{BaO} \cdot \text{B}_2\text{O}_3$, produced by Buckman Labs with the trade name Busan 11-M1, is used as a paint mildewcide. As a result of the concerns surrounding the use of pentachlorophenol, there is a resurgence in interest in copper carbamates. A patent presents a two-step process, involving, first, contacting wood with an aqueous solution of a copper compound, followed by a second immersion, in a carbamate compound, to form an *in situ* treatment (35).

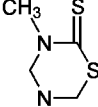
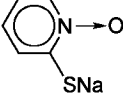
Organosulfur Compounds. These compounds, listed in Table 8, are used in a variety of applications, including cooling water, paint, and metalworking. Methylenebisthiocyanate hydrolyzes rapidly at a pH above 8 to cyanate ion which complexes with ferric iron to poison the cytochrome systems (36).

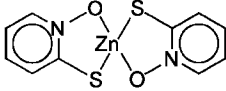
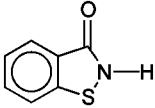
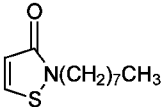
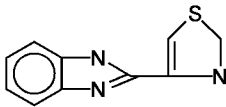
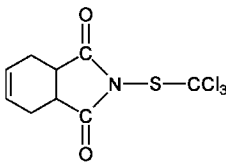
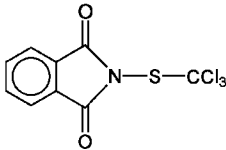
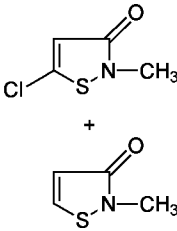
Table 8. Organosulfur Industrial Antimicrobial Agents

Chemical name	CAS Registry Number	Structure	Trade name	Producer	Applications
<i>Bisthiocyanates</i>					
methylenebisthiocyanate (MBT)	[6317-18-6]	NCSCH_2SCN	Cytox 3522	American Cyanamid	paper mill slimicide
vinylenebisthiocyanate	[14150-71-1]	$\text{NCSCH}=\text{CHSCN}$	Biocide N-948 Cytox 3711	Akzo American Cyanamid	cooling water slimicide
chloroethylenebisthiocyanate	[24689-89-2]	$\begin{array}{c} \text{NCSCHCH}_2\text{SCN} \\ \\ \text{Cl} \end{array}$	Cytox 3810	American Cyanamid	secondary oil recovery
<i>Dithiocarbamates</i>					
sodium dimethyldithiocarbamate (Sodam)	[128-04-1]	$\begin{array}{c} \text{S} \\ \\ (\text{CH}_3)_2\text{NCSNa} \end{array}$	Thiostast Thiostop N	Uniroyal	paint mildewcide, slimicide paint mildewcide, slimicide
disodium ethylenebisdithiocarbamate (Nabam)	[142-59-6]	$\begin{array}{c} \text{O} \qquad \qquad \text{O} \\ \qquad \qquad \\ \text{NaSCNHCH}_2\text{CH}_2\text{NHCSNa} \end{array}$	Dithane D 14	Rohm and Haas	cooling water slimicide
zinc dimethyldithiocarbamate (Ziram)	[137-30-4]	$\begin{array}{c} \text{S} \qquad \qquad \text{S} \\ \qquad \qquad \\ (\text{CH}_3)_2\text{NC}-\text{SZnS}-\text{CN}(\text{CH}_3)_2 \end{array}$	Vancide 51Z	RT Vanderbilt	paper mill slimicide
<i>Sulfones</i>					
bis(trichloromethyl) sulfone	[3064-70-8]	$\begin{array}{c} \text{O} \\ \\ \text{Cl}_3\text{C}-\text{S}-\text{CCl}_3 \\ \\ \text{O} \end{array}$	Biocide N-1386	Akzo	paper mill slimicide, cooling water slimicide, secondary oil recovery

Heterocyclics. The heterocyclic compounds used as industrial antimicrobial agents are listed in Table 9. Captan and Folpet are agricultural fungicides that have found some industrial uses, primarily in solvent-based paints. Human toxicology concerns have resulted in some diminishing of markets (37). Lack of solubility has limited the formulation flexibility and product is typically available in a malleable concentrate (44–88° o) (38). A typical synthesis has been disclosed (39).

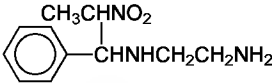
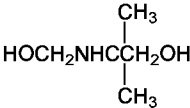
Table 9. Heterocyclic Industrial Antimicrobial Agents

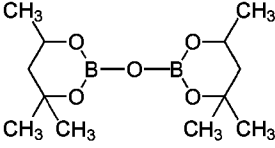
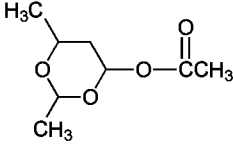
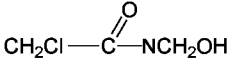
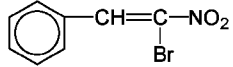
Chemical name	CAS Registry Number	Structure	Trade name	Producer	Applications
<i>Nitrogen–sulfur heterocyclics</i>					
tetrahydro-3,5-dimethyl-2H-1,3,5-thiadiazine-2-thione (DMTT)	[533-74-4]		Metasol D3T Biocide N-521	Calgon Akzo	paper mill slimicide, paint, cooling water slimicide, adhesives
sodium pyridinethione	[3811-73-2]		Sodium Omadine	Olin	metalworking fluids, cosmetics
zinc pyridinethione	[13463-41-7]		Zinc Omadine	Olin	antidandruff agent,

					cosmetics
1,2-benzisothiazoline-3-one	[2634-33-5]		Proxel CRL	Zeneca	latex paint preservatives, pigment slurries, adhesives, metalworking fluids
					
2-(<i>n</i> -octyl)-4-isothiazolin-3-one	[26530-20-1]		Kathon Kathon LP	Rohm and Haas	latex paint mildewcide, leather hides
					
2-(4-thiazolyl)benzimidazole	[148-79-8]		Metasol TK-100	Merck	latex paint mildewcide
					
N-(trichloromethylthio)-4-cyclohexene-1,2-dicarboximide (Captan)	[133-06-2]		Vancide 89RE	RT Vanderbilt	paint
					
N-(trichloromethylthio)-phthalimide (Folpet)	[133-07-3]		Fungitrol 11	H&ls	paint
					
5-chloro-2-methyl-4-isothiazolin-3-one	[26172-55-4]		Kathon 886	Rohm and Haas	paint, metalworking, cooling water, cosmetics
2-methyl-4-isothiazolin-3-one	[2682-20-4]				

Other Nitrogen Compounds. The basis of the sophisticated nitrogen compounds listed in Table 10 is the reaction of formaldehyde with amino compounds. A significant amount of literature details investigation of the mechanism of action, particularly whether or not the antimicrobial activity depends on decomposition to formaldehyde (40–42). These compounds tend to have substantial water solubility and are more effective against bacteria than fungi and yeasts. Key markets for these compounds are metalworking fluids, cosmetics, and in-can preservation of paints (see *ALKANOLAMINES*; *AMINES*, *FATTY AMINES*).

Table 10. Industrial Amine, Alkanolamine, and Nitro-Containing Antimicrobial Agents

Chemical name	CAS Registry Number	Structure	Trade name	Producer	Applications
<i>N</i> -cocotrimethylenediamine	[61791-63-7]	<i>Linear</i> RNHCH ₂ CH ₂ NH ₂		Akzo	secondary oil recovery, cooling water, paper
<i>N</i> -[α-(1-nitroethyl)benzyl]-ethylenediamine	[14762-38-0]		Metasol J-26	Merck	paper mill slimicide
2-(hydroxymethyl)amino-ethanol	[34375-28-5]	HOCH ₂ NHCH ₂ CH ₂ OH	Troysan 174 Nuosept	Troy H&ls	latex paint, resin emulsion
2-(hydroxymethyl)amino-2-methylpropanol	[52299-20-4]		Troysan 192	Troy	adhesives, tape-joint compound
2-hydroxymethyl-2-nitro-1,3-propanediol	[126-11-4]		Tris Nitro	Angus	metalworking,

					
2,6-dimethyl, 1,3-dioxanol-4 acetate	[828-00-2]		Givguard-D XM	Rhône-Poulenc Rorer	textiles
					
1-(3-chloroallyl)-3,5,7-triaza-1-azoniaadamantane chloride	[4080-31-3]		Dowicil 100	Dow	adhesives, floor waxes, latex emulsions, metalworking fluids, paint
2-chloro- <i>N</i> -(hydroxymethyl) acetamide	[6320-16-7]		Grotan HD2	Lehn & Fink	metalworking
					
β-bromo-β-nitrostyrene	[7166-19-0]			Rhône-Poulenc Rorer	cooling water, metalworking
					

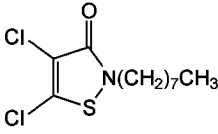
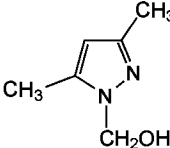
As noted earlier, the most significant trend in the industry has been the decline in the number of available active ingredients due to EPA regulations. Table 12 shows the new active ingredients, together with the year of their obtaining an EPA registration. Bronopol had actually been used as a preservative for cosmetics prior to its registration (45). The importance of these products is not known.

Table 12. Antimicrobials of the 1980s

Chemical	CAS Registry Number	Name	Registration year
methyl-3,5,7-triaz-1-azoniatricyclodecane chloride		BusanR 1024	1987
1-(hydroxymethyl)5,5-dimethyl hydantoin	[116-25-6]	Glyco Serve	1985
2-bromo-2-nitropropanediol	[52-51-7]	Bronopol	1984
methanol, [[2-(dihydro-5-methyl-3 (2 <i>H</i>)-oxazolyl)-1-methylethoxy]ethoxy]methoxyl alkyl(61% C ₁₂ , 23% C ₁₄ , 11% C ₁ , 5% C ₁₈) dimethyl benzyl ammonium chloride	[68391-01-5]	CosanR 145	1983
			1982

Table 13 shows some of the developmental products that have EPA applications pending and may be available in the near future. Sea Nine is a variation on the very successful isothiazolone chemistry. It is claimed to be an improvement over metallic actives used for antifouling paint and wood preservation (46,47). Decylthioethylamine and its water-soluble hydrochloride are claimed to be especially effective at controlling biofilm in cooling water applications (48–50). The hydroxymethylpyrazole shown is also suggested to have properties that are well suited to the protection of aqueous products or emulsions (51,52).

Table 13. Developmental Industrial Antimicrobials

Chemical name	CAS Registry Number	Structure	Producer
2(<i>n</i> -octyl)-4,5-dichloro-isothiozal-3-one ^a	[26530-20-1]		Rohm and Haas
decylthioethylamine ^b	[29873-30-1]	CH ₃ (CH ₂) ₉ SCCH ₂ CH ₂ NH ₂	Dow
decylthioethylamine-hydrochloride ^b	[36362-09-1]	CH ₃ (CH ₂) ₉ SCCH ₂ CH ₂ NH ⁺ ₃ Cl ⁻	Dow
<i>N</i> -hydroxymethyl-3,5-dimethylpyrazol ^e	[85264-33-1]		Buckman

^a Trade name, Sea Nine; used as a marine antifoulant.
^b Trade name, XV-40304-OIL; used for cooling water.

BIBLIOGRAPHY

"Industrial Fungicides" under "Fungicides" in *ECT* 1st ed., Vol. 6, pp. 991–995, by R. G. H. Siu, Quartermaster General Laboratories, U.S. Army; in *ECT* 2nd ed., Vol. 10, pp. 228–236, by N. Joe Turner, Boyce Thompson Institute for Plant Research, Inc.; "Industrial Antimicrobial Agents" in *ECT* 3rd ed., Vol. 13, pp. 223–253, by S. I. Trotz and J. J. Pitts, Olin Corp.

1. J. G. Holt, ed., *Bergey's Manual of Systematic Bacteriology*, 8th ed., Williams & Wilkins Co., Baltimore, Md., 1984.
2. I. E. Akamo, *Fundamentals of Microbiology*, 4th ed., The Benjamin Cummings Publishing Co., Inc., 1994.
3. M. J. Pelczar, Jr. and R. R. Reid, *Microbiology*, McGraw-Hill Book Co., Inc., New York, 1972, pp. 72–73, 412, 424–439.
4. L. S. Goodman and H. Gilman, *The Pharmacological Basis of Therapeutics*, 5th ed., Macmillan Publishing Corp., New York, 1975, pp. 946–959, 987–1017, 1090–1129.
5. S. S. Block, ed., in *Disinfection, Sterilization, and Preservation*, Lea and Febiger, Philadelphia, Pa., 1991, pp. 20–24.
6. O. Shapiro and L. Magier, *Adhes. Age* **34**, 22–24, (1991).
7. U.S. Environmental Protection Agency, subdivision G, *Product Performance NTIS PB-83-153924*, Washington, D.C., Sept. 1982.
8. H. Rossmore, *ASTM Standards on Materials and Environmental Microbiology*, American Society for Testing and Materials, Philadelphia, Pa., 1987.
9. A. A. Aspelin and F. S. Bishop, *Chemicals Registered for the First Time as Pesticidal Active Ingredient Under FIFRA*, Office of Pesticide Programs, U.S. Environmental Protection Agency, Washington, D.C., 1991.
10. *Status of Pesticides in Reregistration and Special Review*, Office of Pesticide Programs, U.S. Environmental Protection Agency, Washington, D.C., 1992.
11. W. Poppendorf, M. Selim, and N. Lewis, "Exposure While Applying Industrial Antimicrobial Pesticides," *Am. Ind. Hygiene Assoc. J.*, in press (1994).
12. Chemical Manufacturers Association, Biocides Panel, Washington, D.C. personal communication, 1991.
13. E. S. Keische and D. Hunter, *Chem. Week*, 24–25 (July 22, 1992).
14. D. L. Price, A. D. Swant, and D. G. Ahern, *J. Ind. Microbiol.* **8**, 83–90 (1991).
15. **U.S. Pat. 5,024,840** (June 18, 1991), L. W. Blakely and co-workers (to Interface, Inc.).
16. Ref. 5, pp. 241–242.
17. J. L. Spier and J. R. Malek, *J. Colloid Interface Sci.* **89**, 68–76 (1982).
18. J. Hanson, *Mod. Paint Coatings*, 50 (1984).
19. Ref. 5, pp. 229–230.
20. Ref. 5, pp. 131–149.
21. J. R. Back, B. R. Friedfeld, and A. A. Boccone, *Biocides—United States*, C. K. Kline, Fairfield, N.J. 1984, p. 29.
22. W. F. McCoy and J. W. Wireman, *J. Ind. Microbiol.* **4**, 403–408 (1989).
23. H. W. Talbot, J. E. Morrow, and R. J. Seidler, *J. Amer. Water Works Ass.* **71**, 349–353 (1979).
24. R. Mullin, *Chem. Week*, 37–40 (May 12, 1993).
25. E. Brantt, *Chem. Eng.* **99**(10) 57,59 (1992).
26. Ref. 5, p. 142.
27. V. S. Brnzal and T. E. Cloete, *J. Ind. Microbiol.* **8**, 273–276 (1991).
28. W. E. Machemer, *Devel. Ind. Microbiol.* **20**, 25–29 (1979).
29. Ref. 5, pp. 334–343.
30. N. M. Rei, T. C. McEntee, and J. F. Brophy, in J. Edenbaum, ed., *Plastics Additives and Modifiers Handbook*, Van Nostrand Reinhold Co., Inc., New York, 1992, pp. 338–355.
31. Ref. 5, pp. 344–357.
32. J. J. Miller, W. E. Brown, and V. J. Krieger, *Devel. Ind. Microbiol.* **22**, 763–770 (1981).
33. R. P. Bratt and co-workers, *Biodeterioration Biodegradation* **29**, 61–73 (1992).
34. M. Dub, *Organo-Metallic Compounds*, Vol. III, Springer-Verlag, New York, 1968.
35. **U.S. Pat. 4,937,143** (June 26, 1990), M. H. West (to Chapman Chemical Co.).
36. J. W. McCoy, *Microbiology of Cooling Water*, Chemical Publishing Co., New York, 1980, pp. 82–83.
37. *Suspended, Cancelled and Restricted Pesticides*, Office of Pesticides and Toxic Substances, U.S. Environmental Protection Agency, Washington, D.C., 1990.
38. J. H. Ayers and D. Lopez, *Chemical Economics Handbook, Fungicides*, SRI International, Palo Alto, Calif., 1984.
39. **U.S. Pat. 2,553,770** (May 22, 1951), A. R. Kittleson (to Standard Oil).
40. H. W. Rossmore and M. Sondossi, *Advances in Applied Microbiology*, Vol. 33, Academic Press, Inc., San Diego, Calif., 1988, pp. 223–274.
41. M. Sondossi, H. W. Rossmore, and W. Wireman, *J. Ind. Microbiol.* **1**, 87–103 (1986).
42. C. P. Barnes and R. G. Eagon, *J. Ind. Microbiol.* **1**, 113–118 (1986).
43. Ref. 5, p. 927.
44. C. Genner and E. C. Hill, in A. H. Rose, ed., *Microbial Biodeterioration*, Academic Press, London, 1981, pp. 276, 277.
45. E. L. Richardson, *Cosmet. Toiletries* **96**, 91, 92 (Mar. 1981).
46. J. C. Harrington, *Comparative Effects of Organic Antifoulant Biocides in Marine Paints*, 29th Marine and Offshore Coatings Conference, Charleston, S.C., June 28, 1989.
47. J. Hall, ed., *American Wood Preservers Association Standards AWP*A, Washington, D.C., 1991.
48. **U.S. Pat. 4,816,061** (Mar. 28, 1989), R. W. Walter, A. G. Relenyi, and R. L. Johnson (to The Dow Chemical Co.).
49. **U.S. Pat. 4,982,004** (Jan. 1, 1992), A. G. Relenyi and C. D. Gartner (to The Dow Chemical Co.).
50. **U.S. Pat. 5,025,038** (June 18, 1991), A. G. Relenyi and C. C. Gartner (to The Dow Chemical Co.).
51. **U.S. Pat. 4,801,362** (Jan. 31, 1989), J. G. Fenyas (to Buckman Lab International).
52. C. G. Hollis and co-workers, in H. W. Rossmore, ed., *Biodeterioration and Biodegradation*, Vol. 8, Elsevier Applied Science, London, 1991.

Thomas McEntee
Morton International, Inc.

INDUSTRIAL HYGIENE

Industrial hygiene is devoted to the anticipation, recognition, evaluation, and control of environmental factors or stresses arising in or from the workplace that may cause sickness, impaired health and well-being, or significant discomfort and inefficiency among workers or among the citizens of the community. It is a profession practiced by over 11,000 industrial hygienists in the United States and many more worldwide. U.S. industrial hygienists are typically members of the American Industrial Hygiene Association (AIHA), which is the largest industrial hygiene organization, the American Conference of Governmental Industrial Hygienists (ACGIH), and the American Academy of Industrial Hygiene (AAIH). Many are certified industrial hygienists (CIH) as a result of meeting the requirements of the American Board of Industrial Hygiene (ABIH). Outside the United States, industrial (also called occupational) hygienists are members of such professional associations as the British Occupational Hygiene Society (BOHS) and the International Occupational Hygiene Association (IOHA).

Industrial hygienists work closely with members of several other professions concerned with workplace health and safety, eg, occupational medicine, occupational health nursing, and safety engineering. All of these groups are involved in the implementation of the laws that regulate workplace health and safety. In the United States the principal law is the Occupational Safety and Health Act (OSHA) (1) enforced by the U.S. Department of Labor (U.S. DOL). Similar laws are in place in almost every country in the world and are proposed by such international organizations as the World Health Organization (WHO) and the International Labor Organization (ILO).

A partial list of the hazards or conditions arising from the workplace (see also PLANT SAFETY) and with which industrial hygienists are concerned includes

<i>Chemical</i>	microwave radiation
carcinogens	
acute poisons	extremely low frequency (ELF) radiation
reproductive hazards	vibration
corrosives	magnetic fields
irritants	ultraviolet radiation
	infrared radiation
	laser radiation
pneumoconiosis producing ducts	
neurotoxins	
nephro (kidney) toxins	<i>Ergonomic</i>
	repetitive strain injury (RSI)
	carpal tunnel syndrome
<i>Physical</i>	back injury
noise	lifting hazards
heat	visual display units
cold	human machine interaction
ionizing radiation	

Industrial hygienists must be able to anticipate what workplace materials or events may give rise to any of these hazards, to recognize the hazards that occur, to evaluate a hazard to determine the degree of risk it presents, and to control hazards so as to reduce risk. The most cost-effective way to deal with workplace hazards is to anticipate them and if possible prevent their occurrence (see HAZARD ANALYSIS AND RISK ASSESSMENT). The industrial hygienist's job begins when a new chemical or process is conceived. Based on data from animal experiments and or human epidemiology relating to a substance or an analogous chemical it is possible to estimate the toxicity of the substance (see TOXICOLOGY). Some substances are benign, others are potent carcinogens, and most are in between. Whenever possible, it is best to avoid using potentially dangerous chemicals. Similarly, potentially hazardous processes that produce excessive noise, heat, or other stress-related situations should be anticipated and avoided. However, the industrial hygienist can usually devise ways to use potentially dangerous chemicals safely (2).

Recognition of Potential Hazards

The process of recognition of potential hazards is based on extensive knowledge of what kinds of hazards may occur in any industry, process, or job activity. Table 1 summarizes the kinds of chemical hazard exposure sources typically found in the chemical process industry. A rating system for each source type determining the degree of exposure to be expected is also given. The recognition process typically proceeds by looking for sources of worker exposure to harmful chemicals and physical agents.

Table 1. Exposure Sources in the Chemical Process Industry

Parameter	Source type ^a	Worker activity	Relative importance	Control ^b
<i>Fugitive emissions or leaks</i>				
pump seal	I or C	no	high	M
flange	C	no	low	M
agitator seal	I or C	no	medium	M
valve stem	C	no	high	M
<i>Process operations</i>				
sampling	I, E	yes	medium	W, E
filter change	I, E	yes	low	W, P
gauging	I, E	maybe	low	E, W
venting and flaring	I or C	no	medium	E
extruding	I or C	yes	medium	E
<i>Material handling</i>				
solid addition	I, E	yes	medium	
liquid transfer	I or C	no	high	E
bagging	C	yes	high	E

drumming	C	yes	high	E, W
bag dumping	I	yes	high	E, W
screening	C	no	medium	E
open mixing	I	no	medium	E, P
banbury mixing	I or C	yes	high	E, P
milling	I or C	no	medium	E, P
Maintenance				
equipment opening	I, E	yes	high	W, P
instrument line draining	I, E	yes	medium	W, P
welding	I, E	yes	high	E, W, P
painting	I, E	yes	medium	W, P
sandblasting	I, E	yes	high	E, P
insulating	I, E	yes	high	W, P ^c
insulation removal ^c	I, E	yes	high	W, P
chemical cleaning	I, E	yes	medium	W, P
degreasing	I, E	yes	low	E
cutting and burning	I, E	yes	medium	W, P
catalyst handling	I, E	yes	high	W, P
Waste handling				
baghouse cleaning	I, E	yes	high	P
drain and sewer venting	I or C	no	high	E
spill clean up	I, E	yes	medium	P
sweeping	I, E	yes	low	W
incineration	I or C	maybe	medium	E
wastewater treatment	C	no	medium	E
sludge handling	I, E	yes	medium	W, P

^a C = continuous; I = intermittent, ie, over discrete intervals of time; E = episodic, ie, nonrandom, the result of an event.

^b M = maintenance, ie, primarily the monitoring and repair of leaks by replacing pump seals, repacking valve glands, tightening flanges, sealing holes in duct work, etc; P = personal protection, ie, the use of an air purifying or supplied air respirator, usually for a short period of time, for a particular hazardous operation; W = work practices, ie, staying upwind of a release source, not spilling volatile liquids on the ground, keeping the work area clean to avoid redispersion of dusty materials; and E = engineering, ie, equipment or process modifications to prevent or contain release such as welded pipe joints, hermetic pumps, vent scrubbers, sealed drains, or local exhaust ventilation.

^c Substitution of less toxic materials for asbestos (qv) is the most common control.

Fugitive Emissions. Fugitive emissions or leaks occur wherever there are breaks in a barrier that maintains containment. Whereas flange and seal leaks are individually small, these can cumulatively amount to the main source of loss from a unit. Even when these emissions are very small and cannot be detected as losses in a material balance, high local concentrations of contaminants can result and lead to overexposure. Furthermore, some leak sources, such as valve stem leaks, tend to gradually increase over time and can become large if not corrected. Other leaks, such as pump seal leaks, which are usually small, can become very large in the event of total seal failure. Overall, in most well-maintained plants, pumps and valves are more important sources of leakage than flanges. For that reason, leak detection and repair efforts usually focus on pumps and valves unless there is reason to suspect flanges. Fugitive emissions, even without catastrophic seal failure, are the origin of a continuous background exposure for workers. Whereas this source of exposure may not, by itself, result in overexposure, its presence reduces the margin within which other exposures may vary and still remain under the accepted limit (see MAINTENANCE).

Process Operations. The operation of a modern chemical plant is typically computer controlled and does not involve any routine operator contact with the feedstock, intermediates, or product (see PROCESS CONTROL). There are, however, a few actions the operators may need to take which can involve contact with process materials. Sampling of process streams is one such task. Whereas use of on-line analyzers has substantially reduced the need for operator-collected samples, the latter are necessary to check the on-line analyzer or wherever on-line analyzers are not used. Exposure during sampling can be very high if the sampling line is flushed by running a quantity of a volatile liquid out on the pad. On the other hand, exposure can be very low where the sample is collected in a bomb from a closed loop. Worker care in following prescribed practices is important.

Many filters in chemical process units are either changed very rarely or are back-flushed automatically so there is hardly any exposure. Some filters, however, require frequent manual changing or cleaning and significant exposure may occur unless operators follow the proper procedure. The filter container should be drained of any toxic material and then flushed and purged as needed so that when it is opened there is only minimal exposure. Zero exposure is difficult to achieve in situations where a disposable paper filter cartridge may retain and slowly release a material that cannot be removed by multiple flushes and purges.

Gauging is often done automatically, but there are occasions where gauging needs to be carried out using a tape dropped through a hatch on the top of a tank. Even where automatic systems are installed, manual gauging may be used as a check. Depending on the nature of the liquid in the tank, vapors can be released more or less actively while the ullage hatch is open. Short of using respiratory protection, the only exposure control applicable to open-hatch gauging is the work practice of standing upwind if the platform at the hatch permits.

Vents and flares are intended to take contaminants released from safety valves away from work areas. However, if an elevated vent is at the level of an occupiable platform on the same or an adjacent unit, a worker may, under certain wind conditions, be subject to the nearly undiluted effluent of a vent. Whereas such elevated platforms may rarely be occupied, a heavy exposure from a vent could incapacitate a worker or cause a fall. Tanks that vent only when being filled are common causes of this concern. The usual solution is to raise the vent above any occupiable platform or, at greater cost, to scrub the vent effluent.

Extrusion is a common way for solid products such as plastics to emerge from closed manufacturing systems. Normally a polymer is hot when extruded and may contain additives and oligomers that are volatile at elevated temperatures. The result is fuming at the extruder head. These fumes can result in employee annoyance, housekeeping problems, and, at worst, depending on composition, health hazards.

In all of the process operations except venting and flaring, exposure is related to worker activity, and to some extent is dependent on worker behavior and the work practices applied. The distinction between those exposures that are impacted by worker behavior and those that, barring the use of respirators, are not is important. The types of control methods to be applied and the methods of exposure measurement to be used are influenced by this difference.

Material Handling. The continuous movement of materials through a process unit does not in itself result in opportunities for release and consequent exposure. However, some material-handling steps are difficult to accomplish with total containment. Wherever quantities of material are allowed to build up and be drawn from tankage or at temporary or permanent storage points, the possibility of release needs to be considered. Liquids entering fixed tanks displace air, which must be vented to avoid overpressuring the vessel. Control of liquid-transfer operations can be achieved by variable volume vessels, such as those having floating roofs, which do not require venting, scrubbing, flaring, or recovering the vented gas stream (see TANKS AND PRESSURE VESSELS).

In drumming and the filling of tank cars and trucks, where the vessel is initially empty, the amount of material being transferred that could be released by displacement depends on how much evaporates during the filling. Rarely does a material evaporate so quickly that the entire volume of displaced gas is saturated. More likely the initial release at the start of filling contains only a small amount and the concentration increases toward saturation

as the filling proceeds. How quickly the concentration in the vented gas increases depends on the temperature and volatility of the material and on the loading mode. Splash loading, where the material leaves the filling spout at the top of the tank and free-falls to the bottom, results in much more evaporation of a volatile material than bottom loading or submerged spout loading where the liquid level rises gently without splashing. Some very volatile liquids and gases that are liquids under pressure are allowed to autorefrigerate by controlled evaporation during transit. Vapors vented from liquid transfer can be collected and sent to an elevated vent or flare, returned to the discharging vessel head space, reliquefied by refrigeration, and returned to storage. One other opportunity for vapor release in material handling is air open mixing. In batch blending or production operations it may be the practice to add liquids or solids to an open vat that already contains some liquid. As the liquid evaporates and as the gas in the head space of the vat is displaced by material addition, some vapor may escape out of the hatch or vat top. The amount of release depends on the temperature, volatility of the liquid, degree of mixer agitation, rate of addition of material, and openness of the vat top or hatch. Control is usually accomplished by local exhaust ventilation systems that maintain a negative pressure in the vat or collect vapors as they escape from the vat, or by closed addition systems such as rotary valves (see MIXING AND BLENDEDING).

Solids handling is often done by open means both because the hazard is perceived to be less and because it is more difficult to design totally closed solids handling systems (see POWDERS, HANDLING). Where solids are handled in closed systems, often as fluids in pneumatic conveying (qv) systems, the potential release problems are similar to those for liquid transfer. Air displaced from hoppers and silos as these are filled may contain dust. Also, the conveying air must be released from the vessel where the conveyed material is deposited. These streams are sometimes sent to an air cleaning system such as a baghouse or other filter or an electrostatic precipitator. Whereas such devices are effective in removing particulates from the air, maintenance of the air cleaner itself sometimes results in a secondary exposure (see AIR POLLUTION CONTROL METHODS).

Semiclosed systems for handling solids often involve the use of big bags or tote bins. In these systems the solid is shipped in a large (>1 t) container which is then lifted into place over a closed hopper or feed mechanism and a sealed connection is made. It is sometimes necessary to rap or agitate the big container to keep the solid moving and this action can result in deterioration or damage to the container or seals with consequent leaks. Also, there is some release of solids when containers are connected or disconnected as well as opportunities for dust generation.

Solids handling in small containers such as bags and drums can be either automated or manual. Automated bag- and drum-filling machines can position the container, fill it using a weighed amount of product, seal it, and label it. Some bag-filling machines exhaust dust generated inside the bag during filling. Dust generation and dispersion during bag and drum filling depend on the dustiness of the solid, the rate and manner of container filling, the degree of care taken in the filling process, the maintenance of machinery, and the cleanup of spills. Bag dumping can be done in open or closed systems. In open bag dumping there are potential releases when bags rupture during handling, when bags are dumped, and when the empty bag is collapsed for disposal. In manual bag handling operations spills from ruptured bags damaged by fork lifts and other traffic can be a significant source of exposure. Blowing dust off machinery and clothing using compressed air adds to the problem. A means of control is frequent cleaning using wet sweeping or vacuuming with heavy-duty installed systems. Automatic bag dumping can be done by machinery that conveys the bag into a closed chamber where it is split open and dumped. The empty bag is then transported to a compactor. The whole system is enclosed and exhausted so that all leaks are inward ones.

Handling of small quantities of solids may be altogether manual. Materials may be scooped out of drums, weighed on scales, and dumped into mixers entirely by hand. Exposure may be insignificant or a large concern depending on the toxicity and dustiness of the material. Some users of small quantities of solid additives may arrange for the supplier to package the material in preweighed batches packed in a bag so that the whole prepackaged container can be added to the mix. Mixing of solids by means of a banbury, muller, calender, or mill is either an open or incompletely enclosed process. Except for the loading of a mixer, there is usually little opportunity for dust created by the escape of powders because the mixing solids are generally wet or tacky. However, mixing of solids often requires considerable energy and generates heat which results in fumes of evaporated and condensed particulate. Local exhaust ventilation is effective in removing these fumes and close manual handling may still result in exposure.

Maintenance. Closed systems contain process materials except for leaks and fugitive emissions or when opened for maintenance. Open system maintenance can add to exposure by disturbing and dispersing deposits of materials in equipment. Most maintenance (qv) is done while the plant is in operation. Thus the maintenance workers are in close proximity to operating equipment for long periods of time. Local contaminant releases and physical hazards such as noise or thermal radiation need to be considered. In addition, the valves or other barriers blocking off the operating parts of the plant may leak into the maintenance work area. There is also the possibility of failure of the barrier. The piece of equipment being maintained should be cleaned as necessary to reduce exposure before it is opened and repaired. Where highly toxic process materials are present, it may be necessary to flush using a low toxicity stream, strip with steam, and purge with nitrogen. Where this is necessary, the equipment design should include the special fittings needed for the flush and purge line connections. Even when cleaning prior to opening is done as completely as possible, it may be necessary to use respirators at least for the initial opening to guard against overexposure resulting from trapped toxic substances. Proper cleaning and opening of equipment lines and vessels where toxic material may be present is complex and requires careful planning and attention to detail in execution.

Turnarounds, or significant periodic overhauls of chemical plant units, are a special case of plant maintenance. Because units are shut down during turnarounds, some risks are avoided, but because the unit is out of production there is also time pressure to complete the work. Contractors and other workers who may not be familiar with the unit may be brought in so that many maintenance activities proceed simultaneously. In this environment, there is the potential for disorganization and mishap resulting in unanticipated releases of chemicals. To conduct a safe turnaround, it is necessary to plan the event carefully in advance. Contingencies should be anticipated to the extent possible and plans made to deal with them. All of the workers involved should be specially instructed in their duties and closely supervised during the entire turnaround from shutdown through startup and back to normal operation. The materials and operations used in maintenance may present a set of hazards quite separate from the hazards of the feedstocks, intermediates, and products of the chemical unit.

Welding. Any of the metals in the rod or the alloy being welded can become airborne in the welding (qv) fume. Zinc and some other metals can cause metal fume fever, a frequent problem for welders. Other metals such as cadmium can produce systemic effects. Chromium, under certain conditions, can be released in the potentially carcinogenic hexavalent form. In addition to these metal fumes, the welding process produces oxides of nitrogen, ozone, and uv radiation. All of the emissions can be controlled by general or local ventilation or by respirators if necessary. The welder's face mask provides only slight respiratory protection. Welding in confined spaces is particularly hazardous owing to the difficulty in delivering clean air to the site of the welding and in exhausting welding fumes.

Painting. Whereas leaded paints are no longer used for domestic painting, these paints are occasionally used in industry. Frequently surfaces being prepared for painting may have remnants of old lead or chrome coatings which could become airborne during scraping or grinding. The solvents used in paints are not highly toxic, but can reach excessive concentrations in poorly ventilated spaces. Low rates of paint application as from brushing produce lower solvent release rates than intermediate rate application by roller or high rate spraying (see COATING PROCESSES). Certain modern coatings (qv), such as polyurethanes and epoxies, present special toxic hazards (see EPOXY RESINS; URETHANE POLYMERS).

Sandblasting. Whereas some modern corrosion-resistant treatments do not require the removal of all rust, sandblasting to clean metal surfaces prior to coating is very common (see METAL SURFACE TREATMENTS). In addition to the metal dust, the very fine fragments broken off from the abrasive particles may be respirable, that is, capable of reaching the deep lung where these may cause damage. The degree of risk depends greatly on the type of abrasive used. Steel balls and walnut shells produce relatively nontoxic dust, as does aluminum oxide. On the other end, fine dust from sand, which is typically composed of silicon dioxide, is very toxic and can produce a serious lung disease. The degree of dust exposure from sandblasting depends on the degree of enclosure and the use of personal protective equipment. Small pieces can be cleaned in fully enclosed blast cabinets having local exhaust ventilation to maintain negative pressure. Large objects, such as truck bodies, which are too large to be done in cabinets, are often cleaned in large booths using down draft local exhaust ventilation. For structures and fixed piping, sandblasting is done out in the open. When blasting in either the booths or in the open, the operator should be protected by a special sandblaster's supplier-air hood. A common problem occurs when the operator uses the hood for physical protection but does not connect the hood to a supply of clean air. When a hood is used in this manner, fine dust can enter the worker's breathing zone under the hood, and the hood does not provide respiratory protection.

Insulation. It is common for workers replacing insulation at older plants to encounter asbestos (see INSULATION, ELECTRIC; INSULATION, THERMAL). The composition of both old and new insulation should be known to be certain that proper procedures are followed. The removal of asbestos-containing insulation is a complex and difficult process requiring personal protective equipment, monitoring, containment, special disposal procedures, stringent work practices, and record keeping (3). Many companies elect to have asbestos removal done by specialized contractors.

Modern nonasbestos insulation frequently incorporates synthetic mineral fibers for strength at high temperatures. These glass or rock wool fibers are usually silicates that are large compared to asbestos, ie, too large to be inhaled into the deep lung. Also, these fibers do not split the long way and thus do not produce as many fine fibers. For these reasons, even if the synthetics are as potent as asbestos fiber for fiber, the synthetics are less hazardous because fewer respirable fibers result. Ceramic fibers, however, are made in respirable size ranges and are, therefore, more hazardous. Also, when ceramic fibers are used inside high temperature furnaces, they may be converted to christobalite which is more toxic than quartz.

Chemical Cleaning. Removal of deposits from inside vessels and pipes is often done using acids, caustics, or strong solvents, the handling of which can cause a number of hazards. Transfers and mixing of small quantities is usually done manually from drums or tank trucks. Application involves pumping materials through hoses or temporary piping. Sometimes strong cleaners are applied as a pressure spray or jet with consequent spattering. The reaction of the chemical with the deposit materials and the metal of the pipe or vessel can produce dangerous gaseous air contaminants. Even when cleaning is done *in situ*, removal and disposal of the cleaner and flushing fluid can cause exposure. Because these operations are infrequent, installed control equipment is rarely used. In most cases workers rely on personal protection equipment plus detailed handling precautions.

Catalyst Handling. A great many chemical manufacturing reactions are made possible by the use of catalysts such as those listed in Table 2 (see CATALYSIS). Catalysts may be divided into two categories: homogenous catalysts, which are dispersed in the reactant mix so that the entire reaction takes place in a single phase, and heterogeneous catalysts, where catalysis occurs at phase interfaces. Homogeneous catalysts are added to reaction streams the same as any other process chemical and are removed by the usual finishing separations process. From the point of view of industrial hygiene, these chemicals are no different than any other additive. Heterogeneous catalysts, on the other hand, are often used in the form of fixed beds, which must periodically be regenerated or removed and replaced (see CATALYSTS, REGENERATION). These latter tasks must be considered as worker exposure opportunities.

Table 2. Catalyst Industrial Hygiene Concerns

Catalyst	Molecular formula	Possible health effects
aluminum oxide	Al ₂ O ₃	nuisance
aluminum chloride	(AlCl ₃) _x	decomposes to HCl; irritation
aluminum alkyls		acute thermal burns from contact, lung damage
chromic oxide	CrO ₃	Cr ³⁺ , low toxicity; may convert to Cr ⁶⁺ , toxic and carcinogenic
cobalt	Co	lung irritation
cobalt hydrocarbonyl	CoH(CO) ₄	acute respiratory failure
ferric oxide	Fe ₂ O ₃	siderosis; low toxicity
molybdenum compounds		pneumoconiosis
nickel compounds		carcinogenic; eg, nickel subsulfide, Ni ₃ S ₂
platinum and compounds		low toxicity; dermatitis
thorium oxide	ThO ₂	low toxicity; radioactive
uranium		kidney damage
vanadium		respiratory irritation

Catalyst charging and topping is an occasional task typically done at the top of a reactor using temporary handling facilities. For this reason local exhaust ventilation is rarely used even when the operation may be dusty and the catalyst toxic. Scrupulous use of personal protective equipment and adherence to work practices is essential to minimize exposure. Respiratory protection can be so critical as to require air-line respirators; skin protection may include full protective suits. When catalysts are dumped from a reactor these may be very dusty because of particle size reduction occurring in the reactor and because of handling. Dumping is often done via chutes, which do a poor job of containing the dust. In addition to protecting the workers, it may be necessary to erect a temporary enclosure to prevent contamination of adjacent work areas. Some catalysts being dumped are pyrophoric. Water sprays used to prevent fires also help control the dust. When catalyst beds are rendered inert, the danger of a release into the work area of large amounts of the inerting gas, which may cause asphyxiation, exists. Most catalyst removal operations are carried out by experienced contractors using special equipment and techniques. It is important that plant personnel not undertake this or any job for which they are not properly trained or equipped. The hazard from what comes out of a reactor may be quite different and much more severe than that from the catalyst that went into the reactor.

Waste Handling. Housekeeping procedures in general can have a significant impact on employee exposure, and certain waste handling procedures can result in very serious exposure if proper precautions are not taken. The best way to keep a plant clean is to not spill in the first place. Management reviews of the origins of spills and accumulated debris not only keep the plant cleaner but prevent loss of valuable material, save cleanup labor, and reduce fire and other safety hazards. Spilled materials in aisles and on walkways can become airborne by redispersion and can be spread onto surfaces and result in skin contact. Dry powders are best cleaned up with either installed or portable industrial vacuum cleaners. Liquid spills can be soaked up using a number of available solvents, and scraped or shoveled into containers. Careful consideration should be given to the methods used to clean floors. Serious worker overexposures have resulted from the use of volatile solvents on large floor areas inside buildings.

Air cleaning systems are often used to remove dust or vapors from plant or process exhaust streams. Dust collecting systems such as filters or electrostatic precipitators that handle heavy loads of dust are usually designed to be self-cleaning, but it is still necessary to enter the air cleaner periodically for inspection or repair. Dust deposits inside the equipment are likely to be stirred up and inhaled by unprotected workers. Baghouses are particularly likely to cause exposure because large amounts of dust may be retained in the cloth and released when the bags are handled.

Wastewater treatment facilities may receive chemical process wastes and spills. These wastes may volatilize on emerging from a closed sewer system into open waste treatment tanks particularly if hot streams have heated the tank. These releases can occur without warning and result in unexpected employee exposure. Covering reduces the hazard and can also reduce air emissions but does require careful design to avoid creating an explosion hazard. Toxic substances trapped in separator or biological oxidation sludge may be released when sludge is filtered, and skin contact can result from sludge handling.

Hazard Evaluation

The evaluation phase of industrial hygiene is the process of making measurements on some set of samples which permits a conclusion about the degrees of hazard. Before conducting an evaluation, it is necessary to make a number of choices of what and where to sample, when to sample, how long to sample, how many samples to take, what sampling and analytical methods to use, what exposure criteria to use in the analysis of the data, and how to report the results. These choices as a whole constitute the evaluation plan. The object is to find if one or more workers have an unacceptable probability of being exposed in excess of some established limit.

Sampling Strategy. A sampling strategy is a careful plan or method to collect exactly those samples which enable required decisions regarding control to be made at the required level of confidence and minimal cost and effort. The basic choices of sampling strategy are where, when, how long, and how many. A detailed discussion of the statistical basis for sampling strategy and the design of sampling programs are covered elsewhere (4–7) (see SAMPLING).

The origin of the complexity of sampling strategy is the great variability of occupational exposure. The concentration of an air contaminant in the space of a workplace varies with time over both short and long periods. Moreover, workers move in varying patterns through an environment where the contaminant concentration varies with location, and the actions of the workers themselves may cause the concentration to vary. All of these sources of variability lead to an exposure distribution which is usually best described statistically by the log normal distribution (5) and which typically has geometric

standard deviations from two to five or more. This means that the upper seventeenth percentile may be as much as from two to five times the mean. This variability is compounded by the problem of estimating the exposure of a group of workers having differing exposures to find the most exposed workers. Compared to this environmental variability, the variability introduced by the sampling and analytical error is small, even for those methods such as asbestos counting, which are relatively imprecise.

Who to Sample. The objective is to find out if one or more employees may be overexposed, then if it is clear who the most exposed employees are, only those employees need be sampled. If their exposure is acceptable, then all those who are less exposed are also within the limits. If the high risk group is overexposed, further evaluation is necessary to find out if anyone else is overexposed. This high risk sampling technique depends on clear knowledge of how exposure is distributed among a group of workers. Lacking that knowledge or lacking confidence in it, it is necessary to discover the exposure distribution among workers by sampling. The exposure of all exposed employees could be measured or some fraction using statistical tables which produce the probability of finding the highest exposure in a population with various numbers of samples.

Where to Sample. Measurements of the concentration of a contaminant in the general air or at a fixed location are often easier than measurements in a breathing zone of a moving worker. Larger, line powered pumps (qv) can be used to collect bigger samples yielding greater sensitivity. Size selective samplers, such as cascade impactors, can be used to give aerodynamic size distribution data. Fixed monitors tied in to computer data acquisition systems are also possible. The problem is that health effect depends on dose, which depends on exposure. The breathing zone measurement, which most closely approximates exposure, is most often not the same as the general air measurement. This is because of the variability of concentration in the space through which the worker moves and the effect of the workers own activities, eg, welding, grinding, smoking, etc, on the concentration. Consequently personal or other breathing zone sampler measurements are needed for comparison to specific exposure criteria. However, area measurements may still be useful for a number of purposes such as control system evaluation.

When to Sample. Smoothly repetitive operations are likely to be homogeneous over time so that the choice of sample period is not likely to bias the result. Less smooth day to day variation and cyclical operations can be accommodated by random sampling. Some experts have found that systematic sampling can be as free of bias as random sampling even in cyclical operations as long as the sampling period does not match the process period (6). An advantage of systematic sampling, in addition to convenience, is that by making use of information from observation, it is possible to decrease the variance of the sample set. Another advantage is that sampling can be directed at high risk events that might be missed by random sampling.

How Long to Sample. The period of the sample should be matched to the period of the exposure criteria. Most standards are referred to as eight hour time weighted averages (TWAs). These standards are for the average exposure over eight hours. Various combinations of individual samples can be used to obtain the equivalent of what would have been measured by one sample of eight hours duration, as shown in Figure 1. When the standard applies to a shorter period, as for example a short-term exposure limit (STEL) which is a 15-min average, samples should be taken to measure over this shorter averaging time. Some limits are supposed to apply to instantaneous concentrations but because there are no truly instantaneous measurement methods (all have some response time) and because peak concentration is known to be a function of averaging time, these limits are somewhat undefined. The best solution when these limits are to be applied is to make a very short (>1 min) period measurement.

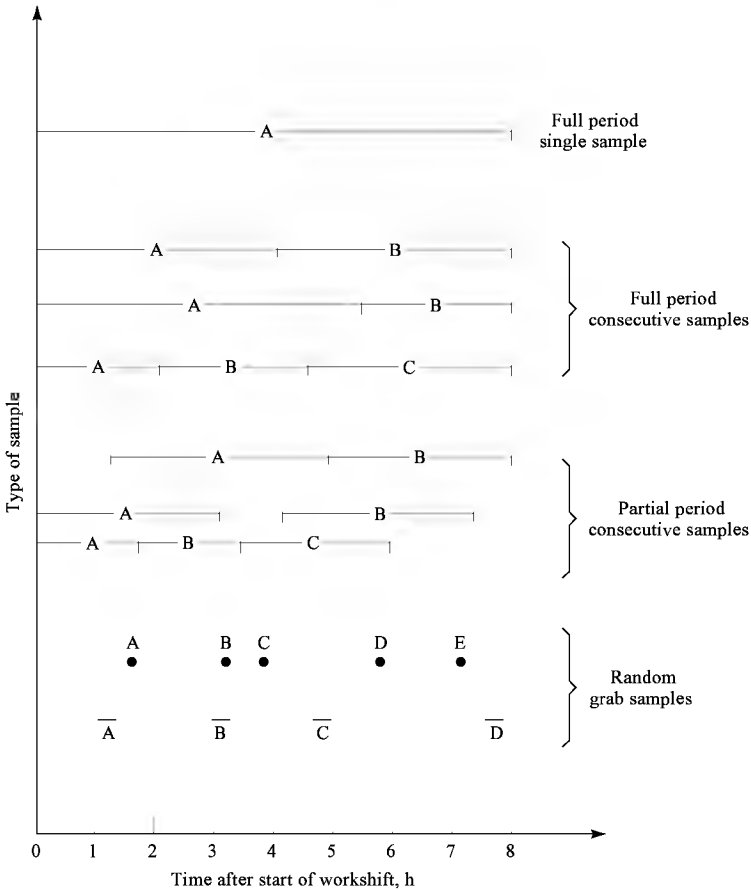


Fig. 1. Sample period options, where each letter, A through E, represents a separate sample.

How Many Samples. A first step in deciding how many samples to collect is to divide what constitutes an overexposure by how much or how often an exposure can go over the exposure criteria limit before it is considered important. Given this quantification of importance it is then possible to calculate, using an assumed variability, how many samples are required to demonstrate just the significance of an important difference if one exists (5). This is the minimum number of samples required for each hypothesis test, but more samples are usually collected. In the usual tolerance limit type of testing where the criteria is not more than some fraction of predicted exceedances at some confidence level, increasing the number of samples does not increase confidence as much as in tests of means. Thus it works out that the incremental benefit above about seven samples is small.

Measurement Method Selection. A measurement method should meet sampling strategy requirements to the degree that the data can be used for decision making. This does not mean that it must be the optimum method with respect to all requirements. The range of methods available is limited and it is often necessary to select a method deficient in one or more attributes but which can yield data from which conclusions can be drawn with the desired degree of confidence. Some of the attributes to be considered in selecting a method follow.

Duration of Sample. When measuring a substance having an 8 h averaging time, a single 8 h sample or several consecutive samples adding up to 8 h is best (see Fig. 1). Short period grab samples are the least satisfactory. For a STEL, the method should be able to collect enough material to provide adequate sensitivity.

Sensitivity. The sampling and analytical method together should ideally have a limit of detection much less than the exposure limit. Less sensitive methods are still usable, however, as long as the limit is easily within the range of the method.

Freedom from Interferences. To avoid spurious results it is necessary that other substances present in the air being sampled do not bias the

result so as to make it unusable. Some error owing to interferences is acceptable if the outside limits of likely error are known and can be taken into account in using the data.

Time to Result. The time required to submit samples to a laboratory, have the samples analyzed, and receive the results is not usually a critical health issue, although promptness in reporting the results of an evaluation adds credibility and impact. On the other hand, some evaluations of acutely acting substance may require immediate results such as a direct on the spot reading.

Intrusiveness. Workers are likely to alter their behavior, consciously or unconsciously, when they are observed. To the extent that a worker's exposure is related to the worker's actions, this change can distort the representativeness of the evaluation. Measurement methods which require the close presence of the person collecting the sample are more likely to influence the result than samples collected with unobtrusive devices worn by the worker.

Proximity to Breathing Zone. Whereas all exposure measurement methods attempt to sample from air that is likely to be inhaled, some methods do so better than others. A sampler fixed some distance away from a breathing area is not usually accurate in measuring exposure. Even using mobile samplers that move with the worker, the few centimeters in distance from the nose and mouth to the position of the sampler, has been found to make a difference.

Accuracy. The more accurate the sampling method the better. Given the very large environmental variability, however, sampling and analytical imprecision is rarely a significant contribution to overall error, or width of confidence limits, of the final result. Even highly imprecise methods, such as dust count methods, do not add much to overall variability when the variability between workers and overtime is considered. An undetected bias, however, is more serious because such bias is not considered by the statistical analysis and can, therefore, result in gross unknown error.

Summary. The technology of air sampling and analysis in the occupational environment makes it possible to take many more samples more conveniently than ever before. Whereas detailed descriptions of specific measuring systems are beyond the scope of this article, the basic systems may be rated. Ratings given in Table 3 are based on the usual or most common systems and devices used in each class. Not all direct reading instruments are insensitive, however, and not all substances can be measured by short period pump–sorbent-type systems. There is considerable variation in the attributes of the specific sampling and analytical system within each class. However, the central or typical performance of various management systems is described to illustrate the use of the selection criteria and to provide benchmarks by which to compare methods.

Table 3. Comparison of Measurement System Attributes

Attributes	Measurement system				
	Direct reading instruments ^a	Continuous monitors	Pumps–sorbent sampler	Detector tubes ^b	Passive badges
sample duration	short	long	short or long	short	long
sensitivity	poor	fair	good	poor	fair
time to result	short	short	long	short	long
intrusiveness	high	low	low	high	very low
breathing zone proximity	fair	poor	good	fair	good

^a Not always specific.
^b Long period tubes available.

Information Gathering. The planning of an evaluation should be complete before any actual measurements are made. The plan should include the sampling strategy element, the choice of sampling and analytical method, and how the data are to be analyzed and tested to arrive at a decision. This last is critical to the planning because weak data cannot support a decision whereas some decisions require no data at all. Once the evaluation plan is set, it should be followed to the extent possible because divergences may bias the result and even compromise the integrity of the conclusion. However, if planning assumptions turn out to be incorrect, it may be necessary to revise the plan. For example, when it is obvious that the sampling and analytical method is not working because of interferences, it would be useless to continue until a new method is found and the strategy altered accordingly. Also, some plans have built-in decision points, such as phased sampling schemes, which decide on a second sequential set of samples based on the results of a first set.

Management and Employee Cooperation. Before beginning to collect data, the cooperation of the managers involved, including the first line supervisor, and of the workers should be secured. Management needs to be informed so that they can be confident that surveillance activities will not upset production or lead to injuries. Workers need to know what the valuation means to them and how the results are to be reported. Everyone needs to know how the measurement is to be conducted so that the actual measurement causes as little disruption as possible.

Sample Integrity. In order to be able to rely on the results of measurements, it is necessary to be sure that the sample as analyzed is the same as it was when collected, and that it is properly identified in the field, in the laboratory, and in the report. Transit times and temperatures should be within the limits allowed for the type of sample and analysis. A series of documents which establish a chain of custody should exist so that it is possible to be sure that the right result goes with the right sample.

Sample Analysis. It is possible to make quality decisions using imprecise methods as long as the imprecision is considered in the sampling statistics. Likewise, it is possible to make good decisions using biased methods if the bias is known and can be offset. In order to properly handle such error sources, it is necessary to know what analytical method is to be used and what its properties are. Further, even when a method is capable of an adequate level of accuracy, it must be demonstrated that the analytical laboratory's results are accurate. To do this, the laboratory must have a quality control program which includes analysis of both internal and external quality control samples. Typically, laboratories having acceptable quality control programs are accredited by the American Industrial Hygiene Association. As an additional check, some spiked samples are usually submitted along with the collected samples and blanks. The usual practice is to subject blanks to all of the handling of a sample (opening cassettes, breaking tube ends), except for drawing the air through the blank.

Factors Influencing Results. Apart from deliberate tampering, which is usually easy to detect, there are other influences which affect sample representativeness and its usefulness in making decisions. The basic question, "What was measured?" must be answered to know if the result can be applied as intended. If, for example, the measurement represents an unusual event, it is probably not useful for characterizing the long-term average exposure of workers. However, it may be useful in deciding if engineering controls are needed to prevent an infrequent but excessive overexposure. If measurements are made in cold weather and all windows are closed, it should be remembered that ventilation is probably better and concentrations lower in the summer, ie, observation can be combined with measurement to understand how to interpret the results.

Decision Process. In many cases, the decision regarding the need for exposure reduction measures is obvious and no formal statistical procedure is necessary. However, as exposure criteria are lowered, and control becomes more difficult, close calls become more common, and a logical decision-making process is needed. A typical process is shown in Figure 2. Even when decision making is easy it is useful to remember the process and the assumptions involved. Based on an evaluation, decisions are made regarding control. The evaluation and decision steps cannot be separated because the conduct of the evaluation, the strategy, measurement method, and data collection are all a part of the decision process.

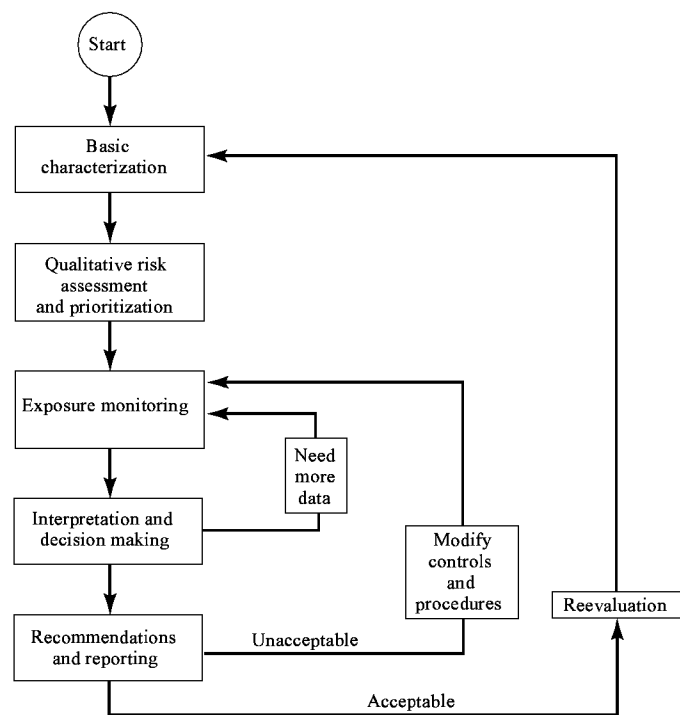


Fig. 2. Decision-making process.

Data Collection. A set of data is collected according to plans using the strategy and methods selected. At the same time, observations are made and recorded which aid in the interpretation of the data.

Data Analysis. First, the raw data must be converted to concentrations over an appropriate time span. When sample periods do not correspond to the averaging time of the exposure limit, some assumptions must be made about unsampled periods. It may be necessary to test the impact of various assumptions on the final decision. Next, some test statistics (confidence limit, etc) (Fig. 3) are calculated and compared to a test criteria to make an inference about a hypotheses.

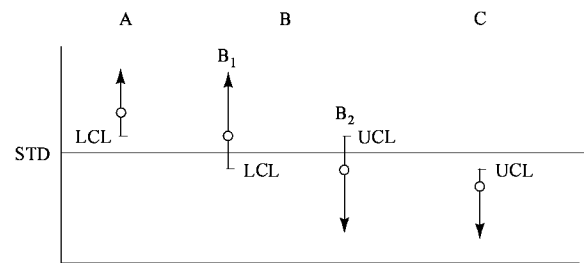


Fig. 3. Confidence limits for exposure levels. A, noncompliance; B, possible overexposure; C, compliance. STD is the standard value, LCL and UCL represent lower and upper confidence levels, between which it is 95% certain that the true exposure lies, and B₁ and B₂ correspond to two separate samples.

Interpretation. Whereas statistical tests establish whether results are or are not different from (over) an exposure criteria, the generality of this outcome must be judged. What did the samples represent? May the outcome, which is inferred to cover both sampled and unsampled periods, be legitimately extrapolated into the future? In other words, is the usual assumption of a stationary mean valid? All of these questions are answered by judgment and experience applied to the observations made at the time of sampling, and the answers are used to interpret the quantitative results.

Conclusion. The quantitative measurements, their interpretation, the calculated statistics, and the exposure criteria all come together to arrive at a conclusion to be drawn with a known chance of being wrong. The data and their interpretation give the extent of the conclusion. The exposure criteria, its origin and basis, define the impact of a conclusion that conditions are unsafe.

Decision. Whereas a conclusion that conditions are to some degree unsafe requires that something be done, what should be done depends on the range and impact of the conclusion. The problem may be easy to correct, or it may no longer exist. The data may describe past conditions that do not presently exist but may recur. It may be that the only possible decision is to undertake significant exposure reduction efforts at great cost. The possibility of each decision should have been anticipated when the evaluation was planned so that the data in hand support the decision that must be made.

Generic Exposure Assessment

In the United States, the Occupational Safety and Health Administration is in the process of developing a generic exposure assessment standard which would apply to chemicals having permissible exposure limits (PELs) listed in Part 1910 of the *Code of Federal Regulations*. Exceptions are the few chemicals for which there are other detailed regulations. This standard is to prescribe how to determine if exposure measurement is required, the frequency of measurement when required, how measurement may be discontinued, what records are to be kept, and how to notify employees. As of this writing, the rule has not been finalized. The evaluation issues raised herein are expected to be covered in such a document. A difficult challenge is how to identify those chemicals which should be measured before any measurements are made, without requiring many measurements to be made of chemicals where exposure is very low, ie, way below the PEL.

Other Agents

Evaluations of occupational exposure to physical agents such as noise, radiation or heat, biological agents, and multiple chemical agents are similar to the process for single chemical substances but have some key differences.

Noise. Technical differences exist between personal noise dosimeters and high accuracy sound level meters and these may alter the usual preference for personal monitors. But it is exposure to noise rather than general room noise that must be estimated for comparison with noise exposure criteria. the logarithmic expression and alternative means of summation (3 vs 5 db doubling) complicate statistics. Exposure criteria for both dose and peak exposure must be evaluated, and space and time variability of noise intensity can be immense.

Radiation. Protection against high voltage and fixed isotope sources of radiation is usually a matter of shielding and the observance of strict

work practices. Evaluation of potential exposure to radiation sources is analogous to safety surveys which look for events and incidents that show weaknesses in procedures. Absenting accidents, exposure should be near background. For some free sources of radiation, such as radon in uranium mines, evaluation is a matter of sampling much as for a chemical substance, except the sampling and analysis can be theoretically and analytically much more complex (see also NUCLEAR REACTORS, 'SAFETY IN NUCLEAR FACILITIES').

Heat. Personal monitoring of the environmental conditions which impose a heat stress on a worker is impractical, so fixed station measurement of such parameters as wet bulb globe temperature are usually made (see TEMPERATURE MEASUREMENTS). These stations are carefully selected so that the results, plus worker location and workload data, can be combined to yield an overall heat stress estimate. Heat strain, the effect on the human, can be estimated from core body temperature, but this is usually only a research tool.

Biological Agents. Evaluation of occupational exposure to biological agents, such as those responsible for anthrax or Legionnaires' disease, is so difficult to do in any quantitative sense that exposure measurement may not be the best pathway to risk estimation. Inhalation is only one route of infection by organisms. Even where inhalation is the primary route, the measurement of exposure is thwarted by the enormous and largely unpredictable variation in exposure. When exposure measurements are made they are difficult to interpret because dose response relationships are not often known, and there are no quantitative standards. Organism variability and human susceptibility make it difficult to predict the consequences of the presence of an organism. Risk assessment is perhaps better based on observation of conditions which could lead to exposure and an observation of biological effect.

Control

The evaluation phase should be planned to yield the data needed to draw accurate conclusions about control needs. The need for certainty depends on how difficult it is to achieve control. If all that is required is a minor change in a work practice or a simple substitution, a reasonable likelihood that the control is necessary may be sufficient. It is much more likely that all the easy controls have already been implemented. Therefore, additional steps involve significant engineering, process, or product changes at considerable cost. In these cases, it is necessary to be very confident of the need to reduce exposure further; ie, it is worth spending considerable effort on data collection to achieve that confidence. There are techniques for calculating the value of information based on the decisions to be made. There is cost associated with being wrong.

Although the evaluation phase comes chronologically between the recognition and control phases, the control options play a considerable role in the extent or intensity of the evaluation phase.

Options. Traditional control options for overexposure are material substitution, process change, containment, enclosure, isolation, source reduction, ventilation, provide personal protection, change work practices, and improve housekeeping. A simple way of looking at selection of control options is to find the cheapest option that results in the desired amount of exposure reduction. It is not actually that simple, however, because the various options differ in ways other than cost and degree of control. Some of the other factors to consider in selection of control options are operability, reliability, and acceptability.

Operability. Hidden costs may result from changes in the way a process operates as a result of a control. For example, enclosure and isolation may diminish the ability of workers to observe the process. Upsets and disruptions resulting from this loss of intelligence are expensive and generate resistance to the use of these controls, no matter how effective.

Reliability. Certain controls are only effective if carefully maintained. Whereas a substitution, if appropriately selected, may need monitoring, a control that depends on a sensor operating an alarm may cease to work after it is installed if it is not carefully checked, calibrated, and repaired. This procedure costs money, time, and supervisory effort, and increases risk.

Acceptability. Personal protection may seem to be the easiest and least expensive way of reducing exposure. Protective clothing such as gloves, aprons, etc, is in fact a necessary adjunct to release control in the prevention of dermal exposure. Respirators are also capable of providing significant protection, but these often have the problem of worker acceptability. Given strong management commitment and supervisor emphasis, it is possible to achieve effective protection using respirators for short periods of use. Long, routine use is almost universally resisted and, as a consequence, actual exposure reduction may not be achieved even when respiratory use is theoretically required. The difficulties of maintaining an effective respirator program are so great that exposure controls which do not rely on worker behavior are easier and more reliable. The same is also true of work practice controls. Wherever difficult, time-consuming work practices are introduced to reduce exposure, there is a tendency to revert to the easy way of doing the job, especially if supervisor emphasis is relaxed.

BIBLIOGRAPHY

"Industrial Hygiene and Toxicology" in *ECT* 1st ed., Vol. 7, pp. 847–870, by C. H. Hine, University of California, and L. Lewis, Industrial and Hygiene Associates; "Industrial Toxicology" in *ECT* 2nd ed., Vol. 11, pp. 595–610, by D. W. Fassett, Eastman Kodak Co.; "Industrial Hygiene and Toxicology" in *ECT* 3rd ed., Vol. 13, pp. 253–277, by G. D. Clayton, Clayton Environmental Consultants, Inc.

1. U.S. Department of Labor, *Occupational Safety and Health Act*, PL 91-596, Washington, D.C., 1970.
2. U.S. Environmental Protection Agency, *Toxic Substance Control Act*. PL 98-80, 99-519, Washington, D.C., 1976.
3. K. Cherry, *Asbestos Engineering, Management, and Control*, Lewis Publishing Co., Boca Raton, Fla., 1988.
4. N. A. Leidel, K. A. Busch, and J. R. Lynch, *Occupational Exposure Sampling Strategy Manual*, publication no. 77-173, U.S. Department of Health and Human Services (NIOSH), Washington, D.C., 1977.
5. N. A. Leidel, and K. A. Busch, in L. J. and L. V. Cralley, eds., *Patty's Industrial Hygiene and Toxicology*, Vol. 3a, 2nd ed., John Wiley & Sons, Inc., New York, 1985.
6. James L. Unmack, "A Comparison of Periodic Versus Random Sampling From an Information Theory Point of View," presented at *CMA Exposure Assessment Workshop*, Washington, D.C., 1986.
7. N. C. Hawkins, S. K. Norwood, and J. C. Rock, *A Strategy for Occupational Exposure Assessment*, American Industrial Hygiene Association, Fairfax, Va., 1991.

General References

American Conference of Governmental Industrial Hygienists, *Advances in Air Sampling*, ACGIH, Cincinnati, Ohio, 1988.

American Conference of Governmental Industrial Hygienists, *Air Sampling Instruments*, 7th ed., ACGIH, Cincinnati, Ohio, 1989.

American Conference of Governmental Industrial Hygienists, *Threshold Limit Values for Chemical Substances and Physical Agents in the Work Environment with Intended Changes for 1993–1994*, ACGIH, 1991, P.O. Box 1937, Cincinnati, Ohio, 45201.

A. Berlin, R. H. Brown, and K. J. Saunders, eds., *Diffusive Sampling, An Alternative Approach to Workplace Air Monitoring*, Royal Society of Chemistry, London, 1987.

W. A. Burgess, *Recognition of Health Hazards in Industry*, Wiley-Interscience, New York, 1981.

H. Checkoway and co-workers, *Am. Ind. Hyg. Assoc. J.* **48**, 515–523, 1987.

Chemical Manufacturers Association, *Improving Air Quality: Guidance for Estimating Fugitive Emissions from Equipment*, 2nd ed., CMA, Washington, D.C., 1989.

Chemical Manufacturers Association, *Responsible Care, Employee Health and Safety Code of Management Practices*, CMA, Washington, D.C., 1992.

L. V. Cralley and L. J. Cralley, *In-Plant Practices for Job Related Health Hazards Control*, Vol. 1, *Production Processes*, John Wiley & Sons, Inc., New York, 1988.

L. V. Cralley and L. J. Cralley, *In-Plant Practices for Job Related Health Hazards Control*, Vol. 2, *Engineering Aspects*, John Wiley & Sons, Inc., New York, 1988.

K. Forsberg and S. Z. Mansdorf, *Quick Selection Guide to Chemical Protective Clothing*, 2nd ed., Global, Denver, Colo., 1994.

E. W. Finucane, *Definitions, Conversions, and Calculations for Occupational Safety and Health Professionals*, Global, Denver, Colo., 1994.

H. H. Faucett and W. S. Wood, *Safety and Accident Prevention in Chemical Operations*, 2nd ed., Wiley-Interscience, New York, 1982.

J. L. Greeno, G. S. Hedstrom, and M. DiBerto, *The Environmental Health and Safety Auditors Handbook*, Arthur D. Little, Cambridge, Mass., 1988.

Health and Safety Executive, *Monitoring Strategies for Toxic Substances*, Guidance Note EH 42, HSE, Bootle, Meseyside, U.K., 1989.

F. E. Jones, *Toxic Organic Vapors in the Workplace*, Global, Denver, Colo., 1994.

R. J. Lewis, *Sax's Dangerous Properties of Industrial Materials*, Van Nostrand Reinhold, New York, 1992.

R. J. Lewis, *Carcinogenically Active Chemicals*, Van Nostrand Reinhold, New York, 1991.

S. Lipton and J. Lynch, *Handbook of Health Hazard Control in the Chemical Process Industry*, John Wiley & Sons, Inc., New York, 1994.

National Institute for Occupational Safety and Health, *Information Profiles on Potential Occupational Hazards*, available from the National Technical Information Service, Springfield, Va., 1993.

National Safety Council, *Personnel Safety in the Chemical and Allied Industries*, NSC, Chicago, 1979.

National Safety Council, *Fundamentals of Industrial Hygiene*, NSC, Chicago, 1988.

R. E. Ney, *Where Did that Chemical Go?*, Van Nostrand Reinhold, New York, 1990.

NIOSH, *Manual of Analytical Methods*, 1984–1991.

H. Raiffa, *Decision Analysis—Introductory Lectures on Choices Under Uncertainty*, Addison-Wesley, Reading, Mass., 1970.

S. M. Rappaport and T. J. Smith, eds. *Exposure Assessment for Epidemiology and Hazard Control*, Lewis, Chelsea, Mich., 1991.

S. Roach, *Health Risks from Hazardous Substances at Work*, Pergamon, Oxford, U.K., 1992.

J. M. Samet and J. D. Spengler, eds. *Indoor Air Pollution*, Johns Hopkins University Press, Baltimore, Md., 1991.

R. A. Wadden and P. A. Scheff, *Engineering Design for the Control of Workplace Hazards*, McGraw-Hill Book Co. Inc., New York, 1987.

Jeremiah Lynch
Exxon Chemical Company

INFORMATION RETRIEVAL

The literature of chemistry and associated fields has increased enormously since 1980. Establishment of subspecialties and newly defined disciplines as well as increased research output have led to an explosion of journals, books, and on-line databases, all of which attempt to capture, record, and disseminate this plethora of knowledge (1). Tertiary reference tools in chemistry and technology (eg, *Kirk-Othmer*, 4th ed.) help track the primary literature. Excellent references that discuss basic chemical information tools are *The Literature Matrix of Chemistry* (1), *Chemical Information Sources* (2), and *How to Find Chemical Information* (3).

Retrieval of chemical information will continue to be an issue of accessing what is available in the fastest, most cost-effective manner. Changes in the ways information is located and retrieved will be driven by technological advances in computer hardware, development of software, and progress in telecommunications. The resources available through Internet are increasing daily. Content of electronic databases has remained basically the same; what has changed are the tools to access the information. Publishers continue to explore the possibilities of electronic media. Ease of information access, for example, through the use of a natural language interface, or the ability to query multilingual databases using a single language, is another emerging issue. The Special Interest Group on Information Retrieval of the Association of Computing Machinery (SIGIR ACM) meets annually to address these and other information issues. Proceedings of their meetings provide an overview of advances in information technology and access. As technology becomes more complex, issues of ownership, copyright protection, reuse of retrieved information, and access costs will need to be examined and resolved (see COPYRIGHTS AND TRADEMARKS).

Libraries and information centers are rapidly moving from purchase and ownership of print and on-line resources to rapid access to these resources; a change in philosophy from "just in case" to "just in time." Libraries are making hard decisions about what to purchase and what to exclude because of the magnitude and cost of a complete collection of available information. These factors have forced severe cutbacks in what is actually bought as well as increases in cooperative purchasing and loan agreements between libraries on the local, regional, national, and even international levels. Increased computer power and technology have made geographic boundaries and limitations obsolete (4). Documents, such as journal articles, can be obtained quickly from other libraries and commercial document delivery vendors (5). Paper copy remains the preferred format for delivery of such documents. Documents can be delivered by regularly scheduled mail, by overnight delivery, or by facsimile transmission. Fax is the method of choice if extremely short delivery time is required; it is relatively inexpensive, ie, generally the cost of a phone call, paper, and supplies, when compared to courier and overnight delivery options. Documents created in an electronic format not only can be easily edited or updated but also can be transmitted rapidly via electronic mail to multiple users. Electronically created files can be loaded onto diskettes for delivery of information if an electronic mail system is not available. Another method of transfer of documents is via modem, which eliminates the physical transfer of diskettes, allows transmission of word-processed, non-ASCII (American Standard Code for Information Interchange) documents with no loss of format, allows transmission of non-English language documents with non-Roman alphabets, and permits rapid, unattended transmission. Speeds in excess of 14,400 bps are currently available.

Although electronic publishing of journals is in its infancy, more and more full text journals are becoming available on-line, allowing printing of a document from the user's computer. The limitation of ASCII format (text only, no graphics) is being addressed by vendors. The American Association for the Advancement of Science's publication *Online Journal of Current Clinical Trials*, which debuted in July 1992, supports text and nontext and publishes a paper within days of acceptance (6). The Research Libraries Group has produced ARIEL, a software package that allows image scanning of a document transmitted through Internet. ARIEL provides images and text of greater resolution than fax and uses standard personal computer (PC) hardware (7).

Location of and access to chemical and technical information other than journal articles is available through computerized information networks. Electronic bulletin board systems (BBS) provide a telecommunications tool to anyone who has a computer and a modem. Questions can be posted and read by thousands of bulletin board users worldwide, and files and software are easily transferred from virtually anywhere to one's computer.

Networks. The rise in popularity and use of Internet has dramatically changed the way information is disseminated. Internet is a worldwide link of thousands of separately administered computer networks of many sizes and types. Each of these networks is connected to as many as tens of thousands of computers; the total number of individual Internet users is in the millions. This high level of conductivity fosters an unparalleled degree of communication, collaboration, resource sharing, and information access. Electronic mail, or e-mail, is a fast, easy, and inexpensive way Internet users around the world can communicate with each other and with users of other independent networks such as CompuServe, Applelink, and the WELL. In the United States, the National Science Foundation Network (NSFNet), a very high speed network that connects key regions across the country, comprises the Internet backbone. The NSFNet will likely evolve into the National Research and Education Network (NREN) as defined in the High Performance Computing Act of 1991 (P. L. 102-194) (8).

Remote login is the ability of a computer user in one location to establish an on-line connection with another computer elsewhere. Once the connection is established, the remote computer is used as if it were a hard-wired terminal of that system. Within the Transmission Control Protocol Internet Protocol (TCP/IP) suite, this facility is called Telnet. Using Telnet, an Internet user can establish connections with a multitude of library catalogues, other bibliographic databases, university information systems, full text databases, data files (eg, statistics, oceanographic data, meteorological data, and geographic data), and other on-line services. Many of these connections are available to any Internet user and can be accessed without an account.

Internet is remarkable in that ease and speed of access are not dependent on proximity. Users can connect to a network on the other side of the globe as easily as, and almost as quickly as, they can connect to a system in the next building. In addition, because many Internet users are not currently being charged according to their level of use, cost seldom inhibits usage. Therefore the barriers of distance, time, and cost, which are often significant when using other forms of electronic communication, are reduced in the Internet environment. Disadvantages include high initial costs for Internet connection and access that requires a computer and telecommunications.

Internet can also be used to transfer files from one computer to another. This function is provided by the File Transfer Protocol (FTP) of the TCP/IP suite. In a method similar to that used with Telnet, an on-line connection is initiated with another Internet computer via FTP. Unlike Telnet, this on-line connection can perform only functions related to locating and transferring files, such as changing directories, listing files, and retrieving files. Every kind of file that can be stored on a computer can be transferred using FTP: text files, software programs, graphic images, sounds, and files formatted for particular software programs, eg, files with word processing formatting instructions. Many computer administrators have set aside archives of files on their machines that anyone on Internet can retrieve. These archive sites support anonymous logins, called anonymous FTP sites, which do not require an individual account to access. To locate files, Internet users can use the Archie service which indexes files from over 900 separate anonymous FTP sites (9).

The three basic Internet applications of remote login, electronic mail, and file transfer are also building blocks of more sophisticated applications that offer increased functionality and ease of network use. Tools such as Gopher, Wide Area Information Servers (WAIS), and World Wide Web (WWW) go beyond the three basic Internet functions to make information on the network easier to locate and use. Detailed descriptions of these tools are available (10). This trend toward more powerful, user-friendly networked information resource access systems should continue as Internet grows and matures.

News groups are a feature of Internet that allow for rapid, worldwide exchange of ideas with others interested in the same field. Most new groups are part of Usenet, the global news service whose topics encompass many areas of science, recreation, and social issues. Users subscribe to new groups of interest, then read and post messages to those groups. The news group "Sci" posts discussions of research marked by special and practical knowledge relating to established scientific disciplines.

Copyright. Any discussion of emerging technologies in storage and retrieval of information leads to a question of copyright compliance. Copyright in the electronic and computer age, using the internal and external networks and document delivery mechanisms outlined above, is a complex and unresolved issue. One of the primary reasons for using document delivery services, other than for obtaining information that is not locally available, is copyright compliance. Document suppliers generally handle payment of any required copyright fees. In the electronic and computer age, establishing who owns a particular piece of information can be difficult, as can determining what can legally be done with information once it is obtained.

Information stored electronically can be both textual and nontextual. It can be linked with other stored files, downloaded from a commercial database, and edited to create entirely new documents, which can be forwarded to others by use of electronic mail. Clearly technology is moving ahead of the law. Traditionally, copyright protection is extended to the original creative expression of an idea fixed in tangible media (11). Because of expanding multimedia use, it is becoming harder to define and apply existing copyright law to protect the original idea.

AT&T Bell Laboratories has built RightPages, a prototype electronic library, as a current awareness alerting tool. This pilot project illustrates many of the copyright problems encountered in the use of advanced technologies. Identified issues include time and cost involved in securing permissions from individual publishers, pricing issues, complexity of licensing agreements and the restrictions they impose, and administrative costs incurred in obtaining and managing all such licensing agreements (12).

The Copyright Clearance Center, which handles licenses for photocopying journals in paper format, has initiated a small electronic copyright pilot project in an attempt to address problems encountered in complying with the law in the computer age (13). Expansion of this initial project is expected. Changes to the traditional copyright law will be market-driven. Publishers and vendors of information who market products in nonpaper formats will need copyright protection while providing an affordable product to make information widely available (14).

Budgeting. These changes in the storage and retrieval of chemical information require that libraries and information centers now consider not only what should be purchased but also what monies should be allocated for the purchase of information in nonprint formats such as CD-ROMs (compact disk read-only memory) and on-line databases. Coupled with this is budgeting for the cost of hardware and software to enable the rapid and cost-effective delivery of needed information (15). The geometric increase in sources, both printed and on-line, has increased the role of information specialist as an expert in the delivery of chemical information. Retrieval from increasingly diverse and complex sources becomes the paramount issue for searchers of chemical literature in the 1990s.

On-Line Database Resources

The on-line information industry has grown dramatically since 1972 when Dialog Information Services, Inc. (Dialog) offered the first publicly available commercial databases (qv). This service, comprising two databases, had fewer than 20 subscribers (16). The 1980s saw incredible growth in the number and range of on-line databases. Databases covering virtually all important subject areas were developed, and significant publications became available via thousands of bibliographic, abstract, textual, directory, and numeric databases. Although databases have existed in the form of bibliographic and legal files since the 1600s, the term database was not coined until the 1950s (17). For electronic databases to be made available publicly, development was required in three primary technologies: computers, communications, and databases themselves (18).

Though unstable, the computer industry grew rapidly during the 1960s, and the final piece of the computer development puzzle, time-sharing, came about late in the decade. Efforts to develop and commercialize time-shared computers were led by General Electric's computer department, which was quickly overtaken by IBM, UNIVAC, and Digital Equipment Corp.

The U.S. government, a primary sponsor of scientific and technological developments that fostered the computer and communications technologies needed by the on-line database industry, also sponsored database development projects, information usage studies, and combined computer database development–usage projects. The successors of some of these projects continue to be prominent and include DIALOG, MEDLINE, BRS, LEXIS, and the Chemical Abstracts Registry System.

During the early 1970s, the necessary telecommunications technology became available with packet switching. ARPANet, the first operational packet-switched digital communications network, was implemented by the U.S. Department of Defense. Commercial systems (eg, Telenet, TYMNET, and GENet) became available shortly thereafter.

In 1981, IBM introduced a low cost PC, which provided avenues for access to on-line databases by end users. In 1986 the president of Dialog noted that, although 85% of DIALOG's customers were information specialists or librarians, 80% of new DIALOG accounts were established for end users (18,19). Users wanted the on-line industry to accommodate their needs and expectations, but the on-line industry did not recognize that the availability of large amounts of on-line information would not, of itself, induce people to use the information.

Database Producers. Producers of databases, also known as database publishers or information providers, determine the content of the databases, produce them, and typically lease or license them to private organizations or database vendors. Database producers may be categorized as government, not-for-profit, commercial industrial, and mixed.

Government examples include the National Library of Medicine (NLM), and National Agricultural Library (NAL). Governments have long been sponsors of scientific and technological developments needed for nurturing the database industry (18). During the 1960s and 1970s, the majority of database producers were government organizations (20). Not-for-profit examples are Chemical Abstracts Service (CAS) and Biosis. Databases produced by academia and professional society-based not-for-profit (NFP) organizations became widely known during the late 1960s and 1970s. Together with government databases, they continue to maintain their value as resources, especially in the sciences. Additionally, some NFP organizations are chartered to disseminate information at little or no cost (21). Some of the commercial databases are founded on government data, often collected by a government agency at substantial cost. They began collecting data to fill the needs of various industries. The United Nations and the European Economic Community offer mixed examples. Many government and nongovernment information sources are coordinating efforts to disseminate information (22).

The percentage of both government not-for-profit databases decreased between 1977 and 1992, whereas the percentage of commercial databases increased. In 1977, 56% of all databases were produced by government agencies, 22% by not-for-profit groups, and 22% by commerce industry. In 1992, 75% of all databases were produced by commerce industry, 15% by government groups, and 9% by not-for-profits. Databases produced by the mixed sector were 11% of the total in 1985 but only 1% in 1992 (see DATABASES).

Database Vendors. Database vendors, sometimes referred to as on-line service providers, often lease or license databases from producers and then add value by putting the information or data into a retrievable form. This is done by processing and preparing the databases for eventual loading onto their computers or time-sharing systems, providing the unique capabilities of the vendor's search software, and making available an audience of potential searchers. Thus database vendors make available to database producers a medium that facilitates speed, searchability, organization, ease of use, full-text searching, use of illustrations and commentary, and links to other pertinent information (20). Additionally, database vendors often provide related services, such as on-line document ordering, selective dissemination of information (SDI), current awareness, and search services, as well as distribution of CD-ROM products to the database users. Some database producers, eg, CAS and NTIS, also act as vendors by creating and operating their own time-sharing systems in order to deliver information directly to customers, thus retaining control over pricing and delivery of their information (20).

Most commercial on-line services did not originally market to end users; their databases were designed from a system's point of view rather than a user's (23). Eventually, the on-line pioneers added menus to facilitate information retrieval. In 1981, Dialog and Bibliographic Retrieval Service (BRS) offered Knowledge Index and BRS After Dark, respectively, but it was not until the mid-1980s that on-line service providers recognized that the use of software to manage information was critical if the industry was to attract customers.

In 1982, the European Space Agency's Information Retrieval Service (ESA IRS) introduced the ZOOM command, providing users with a mechanism to analyze retrieved sets. In 1984, service at a baud rate of 2400 was made available by Tymnet and Telenet for public access to on-line databases. In 1985, the first commercial CD-ROM drives for personal computers became available, along with the first commercial CD-ROM databases. In 1986, Grateful Med software was designed for the National Library of Medicine (NLM). In 1987, Tymnet made service at 9600 baud rate available for accessing public on-line databases, and Dialog introduced OneSearch, a multiple database searching capability (24). Several other systems had multiple-file searching capabilities in place at this time, but Dialog's implementation made a greater impact on the searching community (25). In 1988, Dialog offered image searching and retrieval from the TrademarkScan database, and in 1991 vendors added a host of sources and databases with international coverage in

anticipation of the inception of the European Economic Community in 1992 (26). Despite all these enhancements, since the emergence of on-line databases, difficult command languages and the lack of common commands among the various vendors' systems continue to inhibit growth of end user searching.

Existence of the end user was recognized in the 1980s; it was not until the 1990s that vendors made concerted efforts to accommodate end user needs by providing faster, easier, and more powerful ways to retrieve relevant information. In electronic information retrieval, the term accessibility implies that data exist in electronic form, that data retrieval is cost effective, intuitive, and easy, and that the electronic medium contributes to the quality and usability of the information (27).

Menu interfaces continued to be developed, facilitating access to on-line databases. In late 1989, Dialog introduced HOMEBASE, a menu-driven service designed to guide the user to information about DIALOG services (28). In 1991, the Materials Property Data Network on STN provided menu-driven access; DIMDI (Deutsches Institut für Medizinische Dokumentation und Information) introduced GRIPS-Chem, which accounted for Registry Numbers, synonyms, and other chemical nomenclature when retrieving chemical information from several medical and life science files that did not employ the structure terminology of the predominantly chemical databases; Biosis introduced its Life Science Network whose menu-driven search engine did not require users to either specify which databases were to be searched or to enter search commands; Dialog extended its Corporate Connection menu interface to enable all users to search a large subset of Dialog's databases; Data-Star's Business Focus began providing menu access to several business and marketing files; and Thomson Financial network introduced CORIS, a menu-driven service, to company and industry information sources (27).

Other significant enhancements include the 1991 introductions of DIALOG JOURNAL NAME FINDER, which helped users locate the databases that indexed a particular journal title and the number of records for that title in each file (25), plus two companion files for locating companies and products in DIALOG databases: DIALOG COMPANY NAME FINDER and DIALOG PRODUCT NAME FINDER (29). Orbit's GET command, originally developed by Pergamon InfoLine to analyze patent data, was introduced in 1988; it permitted quick and simple on-line statistical analysis of any printable field in any file within a previously created search set of records and generated output ranked by criteria specified by the searcher (30).

The rapidly changing environment of the on-line industry is reflected in its business transactions. In 1980 BRS was acquired from its founders by a Dutch company, the Thyssen-Bornemisza Group (TBG) (31). In 1983, STN International became available as a commercial on-line database service, created by Chemical Abstracts Service (CAS) to prevent a monopoly by DIALOG in distributing its Chemical Abstracts database (31). In 1986, entrepreneur Robert Maxwell bought Orbit from System Development Corp. In 1988, Dialog was sold by Lockheed Corp. to Knight-Ridder, Inc. In 1989 BRS was sold by TBG to Robert Maxwell (29). In 1991, Ziff Communications acquired Predicasts, Inc. from Information Handling Services Group. In 1993, Data-Star was acquired by Knight-Ridder, Inc., the parent company of Dialog, from Motor-Columbus, and MDL Information Systems Inc., formerly Molecular Design Limited, became a publicly owned company. In 1994 BRS Online Products was acquired by CD Plus Technologies from InfoPro Technologies, and Orbit Online Products was acquired by Questel, also from InfoPro Technologies. In June 1990 Dialog filed a lawsuit against the American Chemical Society (ACS), charging the ACS with trying to monopolize the chemical information business (16). In August 1990 the ACS filed a countersuit, charging Dialog with fraudulent and deceptive accounting procedures, which led to underpayment of royalties (16). In October 1993 ACS and Dialog settled the suit and countersuit, releasing all claims against each other (32).

Advances in technology have led to significant performance improvements in the on-line industry since its birth in the early 1970s. With the lower costs of PCs and modems and with higher modem speeds, ie, 9600 and now 14,400 baud, becoming more common, access to information continues to become more convenient and economical to users. Wireless telecommunications technology facilitates access to electronic information by travelers using portable PCs. Because users are becoming dependent on electronic information, they want to influence database vendors to make interfaces with their systems faster, easier, and more powerful than existing interfaces (33).

As the end user population continues to expand, the issue of cost of on-line access is moving to the forefront. The end user market has been nurtured by on-line instructional programs provided by vendors at reduced prices for educational institutions. The skilled users from these programs expect database vendors and producers to review their pricing algorithms. Alternative pricing strategies based primarily on connect time have been implemented by most vendors and fixed price subscription services have become readily available to end users. Primary database vendors that offer or are developing fixed price options include Mead Data Central, Dow Jones News Retrieval and DataTimes, NewsNet, Dialog, and OCLC (34). The environment of the 1990s is one in which database vendors and producers are competing for survival, and end users are leading the way to what could represent significant growth in the on-line database industry.

BRS Online Products. This vendor comprises BRS Online Service, BRS Colleague, BRS After Dark, and BRS Morning Search (35). The strength of BRS is in medical, physical, and social sciences as well as business and news databases of value to the health care and pharmaceutical industries (36). BRS Online Service contains over 150 bibliographic and full-text databases in the areas of biomedicine, science, technology, business, economics, humanities, and social sciences (37). BRS After Dark is an after-hours PC oriented version of BRS Online Service, offered at reduced rates (37). BRS Morning Search is available only in Europe and retrieves information from the BRS Online Service databank. BRS Colleague provides access to the BRS Online Service databank, but it is a menu-driven on-line service designed for use by health professionals with or without on-line search experience (37).

Cambridge Crystallographic Data Centre. CCDC focuses on collecting, evaluating, and disseminating crystal and molecular structure data obtained by diffraction methods. CCDC maintains the Cambridge Structural Databases, which contain about 110,000 bibliographic, chemical connectivity, and numeric data entries (37) for crystal structures of organic and organometallic compounds analyzed by x-ray or neutron diffraction methods and reported since 1935.

Chemical Information Systems. CIS is a collection of approximately 30 publicly accessible on-line databases of numeric data, bibliographic references, and some full text. The databases contain information on specific chemical substances, including toxicological and carcinogenic research data, hazardous materials handling, chemical and physical properties, safety and health effects, and spectroscopic and pharmaceutical data. Also accessible on Chemical Information Systems (CIS) are databases that provide regulatory information on events or actions at specific sites. The system was originally developed during the 1970s and 1980s under contract to the U.S. Environmental Protection Agency (EPA) and National Institutes of Health (NIH) and has been made available since 1984 to the public from CIS, now a division of PSI International, Inc. (38).

Data-Star. This is Europe's leading on-line database service (39) and covers worldwide business news, financial information, market research, trade statistics, business analysis, healthcare pharmaceuticals, chemicals petrochemicals, chemical industry, biomedicine life science, biotechnology, and technology, with an emphasis on Europe. It was originally formed as a joint venture among BRS, Predicasts, and Radio Suisse (the Swiss telecommunications company) (37). Data-Star offers access to about 300 bibliographic, abstract, directory, and full-text on-line databases, of which approximately 150 are also available on Dialog (40).

DIALOG Information Retrieval Service. DIALOG focuses on business, scientific, technical, and professional information including chemistry, current events, economics, engineering law, medicine, and social sciences, and provides access to over 400 bibliographic, numeric, full-text, and directory databases. All DIALOG databases are searchable by commands, and most have menus (41). DIALOG also includes most of the full-text newspapers formerly offered by VU TEXT Information Services, Inc., a division of Dialog that closed in December 1992. DIALOG is the oldest scientific technical on-line service and the world's largest on-line information retrieval service.

ESA-IRS. The European Space Agency's Information Retrieval Service covers chemistry, aerospace astrophysics, agriculture food science, biomedicine, physics, health and safety, data processing, earth and environmental sciences, education research, electronics, energy, management science, metallurgy, remote sensing, finance, and news. ESA-IRS offers access to over 200 bibliographic and factual scientific and technical databases. This European on-line database vendor was established in 1966 to meet the needs of the ESA, which promotes cooperation among European states in space research and technology (27).

MDL Information Systems, Inc. MDL provides modular software systems for managing chemical information, as well as related molecular and reaction databases for use with the software. MDL's database management programs, MACCS-II and REACCS, provide access to compound and reaction databases and also have the capability to manage user-created databases (37). Although MDL is not considered to be an on-line database vendor, it is mentioned here because of the value of its information products and services to the chemical industry.

Mead Data Central. MDC provides access to electronic legal, news, and medical information via its LEXIS, NEXIS, and MEDIS services, respectively. LEXIS, available since 1973, contains archives of federal and state case law, codes, and regulations, as well as specialized libraries covering various fields of legal practice, such as tax, securities, banking, environmental, and insurance. NEXIS, available since 1979, is a news and business information service providing on-line access to over 750 full-text and 2000 abstract sources. LEXPAT, accessible via NEXIS, contains the complete text of patent and trademark information for U.S. patents issued since 1975. MEDIS contains bibliographic and full-text entries of over 40 current medical journals and textbooks.

MEDLARS. The U.S. National Library of Medicine's (NLM) Medical Literature Analysis and Retrieval System contains over 30 bibliographic databases covering medicine, dentistry, nursing, pharmacology, toxicology, cancer, veterinary medicine, and allied health professions (37). The MEDLARS electronic storage and retrieval system was established at NLM to provide bibliographic access to NLM's biomedical literature collection.

ORBIT Online Service. Orbit is known for its coverage of patents, chemistry, engineering, and occupational health and safety (36). It provides access to about 100 bibliographic, numeric, full-text, and directory databases (42). ORBIT became available in 1972 as the second commercial provider of on-line information, after Dialog. In 1992, Maxwell Online, the parent company of ORBIT Search Service and BRS Information Technologies, changed its name to InfoPro Technologies. It also refocused its two on-line divisions: ORBIT Online Products, featuring ORBIT Online Service, is concentrating on providing patent and patent-related information; BRS Online Products is concentrating on providing comprehensive medical information. In 1994, ORBIT Online Products was acquired by, and became a division of, Questel (a subsidiary of the France Telecom Group). Questel has since changed its name to Orbit-Questel, Inc.

Questel. This is a division of Orbit-Questel, Inc. and a subsidiary of France Telecom (37) and provides on-line access to over 70 bibliographic, abstract, full-text, and structure databases focusing on science technology, chemistry, European business and news, patents, and trademarks (43). Questel provides two on-line search systems for retrieving chemical structures in databases: Generic DARC, the integrated chemical structure search system, which operates on individual specific compound databases, and Markush DARC, which is designed for inputting, storing, and retrieving compounds included in the definition of generic structure representations commonly used in patents (37) (see PATENTS AND TRADE SECRETS).

STN International. The Scientific and Technical Information Network provides access to approximately 160 bibliographic, chemical structure, numeric, reaction, full-text, and directory databases. Topics addressed include chemistry, engineering, health and safety, math, physics, geology, biotechnology, medicine, energy, materials science, pharmacology, and government regulations (44). STN International is operated worldwide by three nonprofit organizations: Chemical Abstracts Service in North America, a division of the American Chemical Society (ACS), the Japan Information Center of Science and Technology in Tokyo, and FIZ Karlsruhe in Germany.

Chemical Information and Search Methods and Services

Chemical information is reported and recorded in many forms, and a wide variety of databases have evolved to collect the various types of information. The following tables outline the bibliographic, business, structure, numeric, spectra, and reaction databases currently available; their producers and vendors; and the subject matter they cover.

Bibliographic/Technical. The principal databases in which bibliographic chemical information is stored are listed in Table 1. Examples of the use of these databases include searching the CA file to find the published work of a certain author, or the World Textiles file to determine the extent of weft knitting machinery use in Europe.

Table 1. Bibliographic/Technical Databases

Database	Producer	Vendor	Subject coverage
Agrochemical Handbook	Royal Society of Chemistry (RSC)	Dialog	information related to ingredients used for pest control
Analytical Abstracts (AA)	RSC	Dialog, Orbit, STN	literature on analytical chemistry
APILIT	American Petroleum Institute	Dialog, Orbit, STN	petroleum energy industry
APIPAT	American Petroleum Institute	Dialog, Orbit, STN	patents of interest to the petrochemical industry
BIOSIS Previews	Biosis	BRS, Dialog, ESA-IRS, Data-Star, Orbit, STN	life sciences
CAFile	Chemical Abstracts Service (CAS)	STN	chemistry and chemical engineering abstracts
CA Registry File	CAS	Orbit (dictionary), Dialog (dictionary), Questel (dictionary and structure), STN dictionary and structure)	chemical substance information and identification
CA Search	CAS	BRS, Data-Star, Dialog, ESA-IRS, Orbit, Questel, STN	chemistry and chemical engineering citations
CAB ABSTRACTS	CAB International	BRS, Data-Star, Dialog, ESA-IRS, STN	agricultural science and related areas of biology
Ceramic Abstracts	American Ceramic Society	Orbit, Dialog, STN	ceramics
Chemical Engineering and Biotechnology Abstract (CEBA)	RSC	Dialog, Orbit, STN	plant and process chemical engineering
Chemical Journals of the American Chemical Society (CJACS)	American Chemical Society (ACS)	STN	full-text journals published by ACS
Chemical Journals of the Royal Society of Chemistry	RSC	STN	full-text journals published by RSC
Chemical Safety NewsBase (CSNB)	RSC	Dialog, ESA-IRS, Orbit, STN	health and safety in chemical industry
Chinese Patent Abstracts in English Database	European Patent Office (EPO)	Dialog, Orbit	Chinese patent (English abstracts)
CLAIMS	IFI Plenum Data Corp.	Dialog, Orbit, Questel, STN	U.S. patents
COMPENDEX PLUS	Engineering Information, Inc.	Data-Star, Dialog, Orbit, STN	engineering
CORROSION	InfoPro Technologies	Orbit	corrosion
Current Biotechnology Abstracts (CBA)	RSC	Data-Star, Dialog, ESA-IRS	biotechnology, including legal and safety issues
Current Patents	Current Patents Ltd.	Data-Star	pharmaceutical patents (U.S., British, and European)
Evaluation Fast-Alert			
Dissertation Abstracts Online	University Microfilms International (UMI)	BRS, OCLC, Dialog, Data-Star, STN	doctoral dissertation (accredited North American universities)
EMBASE	Elsevier Science Publishers	BRS, Data-Star, Dialog, STN	biomedicine, human medicine
Energy Science & Technology	U.S. Department of Energy	Dialog, STN	energy
EPAT	France Institut National de la	Questel	patents applied for and published in the

European Directory of Agrochemical Products	Propriete Industrielle (INPI) RSC	Dialog	European Patent Office European agrochemical products
GenBank	National Institute of Health	STN	bio-sequences, DNA RNA sequences
INPADOC	European Patent Office	Dialog, Orbit, STN	patent family and legal status
INSPEC	The Institution of Electrical Engineers (IEE)	Dialog, STN	literature on physics, electrical engineering and electronics, control theory and technology, computers and computing
JAPIO	Japan Patent Information Organization	Dialog, Orbit, Questel	patents abstracts of Japan
LEXPAT	Mead Data Central, Inc.	MDC	full-text patents
MEDLINE	U.S. National Library of Medicine	BRS, Data-Star, Dialog, Medlars, STN, Questel	medicine, life sciences
NEWCRYST	FIZ Karlsruhe	STN	inorganic and organic crystal structures
NTIS Bibliographic Database	U.S. National Technical Information Services	BRS, Data-Star, Dialog, ESA-IRS, Orbit, STN	government research
Paperchem	Institute of Paper Science and Technology	Dialog	pulp, paper, board
PATDD	Deutsches Patentamt	STN	citations and abstracts from former German Democratic Republic
PATDPA	Deutsches Patentamt	STN	reference, abstract, and illustration patents published by the Deutsches Patentamt
PATOSDE	Wila Verlag Wilhlm Lampl GmH	STN	citations to patents from the German Patent Office (Deutsches Patentamt)
PATOSEP	Wila Verlag Wilhlm Lampl GmH	STN	citations to patents granted by the EPO
PATOSWO	Wila Verlag Wilhlm Lampl GmH	STN	citations to patents published by the World Intellectual Property Organization (WPIO)
PHARMSEARCH	INPI	Questel	patents' citation to French, European, and U.S. pharmaceutical patents
Pira Abstracts	Pira International	Data-Star, Dialog, Orbit, STN	paper, pulps
Rapra Abstracts	Rapra Technology Ltd.	Data-Star, Dialog, Orbit, STN	rubber and plastics industries
SciSearch	ISI	Data-Star, Dialog	science, biobusiness technology
Thomas Register Online	Thomas Publishing Online	Dialog	North American companies and their products
U.S. Patents Fulltext	U.S. Patent and Trademark Office (USPTO)	Dialog	full-text U.S. patents
World Patent Index (WPI)	Derwent Publications, Ltd.	Dialog, Orbit, Questel, STN	chemical, electrical, mechanical patents
World Surface Coating	Paint Research Associate	Orbit	coating, paints
World Textiles	Elsevier Science Publishers	Dialog	fibers, fabrics

Business/Industrial. The principal databases in which business and industrial information is stored are listed in Table 2. Examples include finding a phone number for a company in Wyoming involved in health care or determining the potential market for a new herbicide in the Far East.

Table 2. Business/Industry Databases

Database	Producer	Vendor	Subject coverage
ABI INFORM	UMI Data Courier Management	BRS, Data-Star, Dialog, ESA-IRS, MDC, Orbit, STN	new product, business management, electronic data processing
BIOBUSINESS	Biosis	BRS, Data-Star, Dialog, STN	biomedical research
BUSINESSWIRE	Business Wire	Dialog, Dowjones, MDC, Newsnet	industrial news, press releases
CENDATA	U.S. Census Bureau	Compu-Serve, Dialog	census data
Chemical Industry Notes (CIN)	CAS	Dialog, Orbit, STN	chemical business news
Commerce Business Daily	U.S. Department of Commerce	Dialog, Newsnet	government services, defense contracts
Conference Papers Index	Cambridge Scientific Abstracts	Dialog, STN	technical papers, conference papers
Dialog Journal Name Finder	Dialog Information Services, Inc.	Dialog	Dialog search aid
Dialog Product Name Finder	Dialog Information Services, Inc.	Dialog	Dialog search aid
DISCLOSURE DATABASE	Disclosure Inc.	BRS, Data-Star, Dialog, Dowjones, Lexis	financial statement
Dun's Market Identifiers	Dun & Bradstreet Information Services	Data-Star, Dialog, DowJones	corporate data
ERIC	U.S. Department of Education	BRS, Data-Star, Dialog	education programs and projects
Federal Register	U.S. Printing Office	Dialog, MDC, Lexis	U.S. government rules and regulations, premanufacturing notices
Food Science and Technology Abstracts	International Food Information Services	Data-Star, Dialog, Orbit, STN	food science
Harvard Business Review	John Wiley & Sons, Inc.	BRS, Data-Star, Dialog, MDC	general management, business reviews
Health Periodicals Database	Information Access Company	BRS, Data-Star, Dialog, Compuserve	health, nutrition, and fitness
Health Planning and Administration	U.S. National Library of Medicine	BRS, Data-Star, Dialog, Medlars	health care delivery
ICC International Business Research	ICC Stockbroker Research Ltd.	Dialog	industry and stock reports
International Pharmaceutical Abstracts (IPA)	American Society of Hospital Pharmacists	BRS, Data-Star, Dialog	pharmacy literature drugs
INVESTEXT	Thomson Financial Networks	Compuserve, Data-Star, Dialog, DowJones, MDC, MDL, Newsnet	investment analyst report
MANAGEMENT CONTENTS	Information Access Company	BRS, Data-Star, Dialog	business and management topics

Marquis Who's Who (MWW)	Marquis Who's Who, Inc.	Dialog	notable Americans
MATERIALS BUSINESS FILE	Materials Information	Data-Star, Dialog, Orbit, STN	ceramics, iron, composites, and steels
NTIS Bibliographic Database	U.S. National Technical Information Services	BRS, Data-Star, Dialog, ESA-IRS, Orbit, STN	government research, reports, and projects
Nursing and Allied Health	CINAHL Information Systems	BRS, Data-Star, Dialog	nursing, health professions
PHARMACEUTICAL NEWS INDEX (PNI)	UMI Data Courier, Inc.	BRS, Dialog, Orbit, STN	pharmaceuticals, health
Pharmaceutical Business News	FT Business Enterprises Ltd.	MDC	new pharmaceuticals
Pharmaprojects	PJB Publications Ltd.	BRS, Dialog, STN	drug development and licensing
Pollution Abstracts	Cambridge Scientific Abstracts	Dialog	pollution
PR Newswire	PR Newswire Association, Inc.	Dialog, MDC	business and financial news
PTS News	Predicasts	Dialog	new products
PTS PROMPT	Predicast	Data-Star, Dialog	marketing, new technology
SEC Online	SEC Online, Inc.	Dialog, MDC	companies on U.S. Securities and Exchange Commission
Textile Technology Digest	Institute of Textile Technology	Dialog	textiles
Textline	Reuters Limited	Dialog	European news
THOMAS REGISTER ONLINE	Thomas Publishing Co.	Dialog	information manufacturers
TOXLINE	U.S. National Library of Medicine	BRS, Dialog, MDC	drugs, toxicology
Trade and Industry	Information Access Company	Dialog	trade news
World Textiles	Elsevier Science Publishers	Dialog	textiles

Structure. Structure searching involves matching a query compound against a machine-readable file of chemical structures. Structure searching determines if a compound is present in a file and retrieves it along with any associated information (45,46). Chemical structure files are compiled as novel chemicals, and compounds are registered and given unique identifiers, eg, the CAS Registry Number, which is assigned sequentially to each new structure entering the system. Such identifiers link the structure with all related information in a system, such as chemical names and bibliographic data.

Substructure searching involves retrieval of all the compounds in a file containing some specified portion of a chemical structure, irrespective of the rest of the molecule in which the query substructure occurs (47). Substructure searching is much more complex than structure searching because of the incomplete specification of attachments to or within the substructure (48,49). A two-stage process is normally used for substructure searching. First a file is rapidly scanned for characteristics of the query structure; this eliminates all structures that cannot match the query and produces a subset of query-matching structures. Second, the query substructure is compared to the subset of the file that passed the initial screening to determine if the query substructure is present.

Beilstein File. Beilstein first went on-line on STN International in December 1988 with information on 350,000 compounds, and several months later it was also on Dialog. Beilstein Online now comprises data on five million compounds. The organic substance records contain the critically reviewed and evaluated documents from the *Beilstein Handbook of Organic Chemistry*, main volume and supplements 1–5, which cover the chemical literature from 1779 through 1979. These evaluated data are indicated as Handbook Data in the notes of literature references. The Beilstein File also contains organic substance records for unreviewed excerpts from the primary literature from 1980 to 1991.

The records in the Beilstein File contain structural data and corresponding structural images, numeric data for chemical and physical properties, Beilstein Registry Numbers, CAS Registry Numbers for most substances, and bibliographic data for references to the primary literature. All information is searchable with the exception of stop-words. The stop-words are defined as articles and prepositions (an, by, for, from, of, the, to, and, with) and other frequently occurring words which do not form useful indexing entries and are not directly searchable in an on-line system. The database is in English, but text descriptions of the following fields are in German: Biological Function (BF), Crystal Property Description (CPD), Ecological Data (ECOL), Isolation from Natural Product (INP), Purification (PUR), Toxicity (TOX), Use of Compound (USC), and most Notes (NTE). The Basic Index contains the Beilstein Registry Number, CAS Registry Number, molecular formula, and single words from selected fields.

Structure searching and display software are host-specific. The Softon Substructure Search System (S4) was developed by the Beilstein Institute and Softon of Graefelfing Germany (50). It is a full structure and substructure searching module. The S4 is used in-house by the Beilstein Institute and is operated by DIALOG. STN uses CAS ONLINE's messenger software for on-line structure searching of the Beilstein on-line database (51).

Gmelin File. This file became available on STN International in December 1991, and is comprised of information on 277,458 compounds, a number that is expected to increase to about 700,000 by 1997. The inorganic substance records contain the critically reviewed and evaluated documents from the *Gmelin Handbook of Inorganic and Organometallic Chemistry*, main volume and supplements, covering the chemical literature for the period 1817–1970. Also included are selected data from a pool of 112 journals of inorganic, physical, and organometallic chemistry plus other journals of physics from 1988 to the present. The records in the Gmelin file contain structural data and corresponding structural images, numeric data for chemical and physical properties, Gmelin Registry Numbers, CAS Registry Numbers for most substances, and bibliographic data with references to the primary literature. The database language is English. The Basic Index contains the Gmelin Registry Number, the CAS Registry Number, molecular formula, and single words from CT (control term) and CTM (control term to multicomponents system) (52).

Registry File. This CAS file contains more than 11.8 million chemical substance records. About 8,000–14,000 records are added each week as new substances are identified by the CAS Registry System. The substance records contain CAS Registry Numbers, chemical names, structures, molecular formulas, ring data biosequence information, and classes for polymers. All of this information may be displayed.

Substance information in the Registry file may be searched in a variety of ways, eg, structure information may be searched using structures built on-line with the Structure command or with codes (screen numbers) for predefined structural fragments or class identifiers. Another option is to upload structures drawn using STN Express or other structure-building software, eg, ChemDraw. Protein and nucleic acid sequences may be searched using codes for amino acids or nucleotides. Substance names may be searched using complete names or name fragments. Complete molecular formulas, molecular formula fragments, and information derived from these formulas, including element counts, atom counts, formula weights, and element symbols, may be used to retrieve compounds from the file. Ring identifiers and ring analysis terms may be used to retrieve substances containing ring systems. Polymers may be retrieved using polymer class terms. Alloys may be retrieved using weight percentages and relative compositions.

Answers from all of these searches contain CAS Registry Numbers. Answer sets may be combined, using the Boolean operators AND, OR, or NOT, with other answer sets or with text terms, such as names or molecular formulas. Any answer set also may be used to define subsets of the file for subsequent structure searching. Answer sets of up to 10,000 Registry Numbers from any type of search in this file may also be used as search terms in other files, such as the CA or CAOLD files (53).

CAS/STN International. CAS STN offers structure searchable files such as Registry, Beilstein, MARPAT, CASREACT, and Gmelin; a variety of learning files, eg, LRegistry, LBeilstein, LMARPAT, and LCASREACT; and software products such as STN Express for on-line structure and substructure searching. Chemical Abstracts Service, a division of the American Chemical Society, has published *Chemical Abstracts* since 1907 and jointly operates STN International with FIZ Karlsruhe and the Japan Information Center of Science and Technology.

A number of files under the generic title CAS ONLINE are available on-line on STN International. The system software, MESSENGER, includes chemical substructure, text, and numeric data searching facilities. Chemical structures and Registry Numbers are contained in the CA Registry file. The four ways to search the structures are EXA, FAM, SSS, and CSS.

EXA (exact) search retrieves the input structure and its stereoisomers, homopolymers, ions, radicals, and isotopically labeled compounds. FAM (family) search retrieves the same structures as EXA, plus multicomponent compounds, copolymers, addition compounds, mixtures, and salts. SSS (substructure) search uses a range of possible substituents and bonds in the input structure. CSS (closed substructure) search is a more restrictive

substructure search with limitation on allowed substitution. Structures can be input textually, using alphanumeric characters, or graphically, using a graphics terminal or personal computer.

A number of graphics front-end packages can be used. STN Express, marketed by STN International, is a completely integrated software package that enables the user to perform on-line structure and substructure searches. This sophisticated software program allows the searcher to draw actual chemical structures to be used in a query on STN. Logon procedures as well as structure uploads to STN are automated. Structure drawing is done off-line, allowing the user to build the structure without incurring connect charges. A sample search of 5% of the file can be run at no cost, other than connect charges. This projects whether the search will succeed or whether it is too broad. Once a Registry file search is complete, it is possible to switch to the CA file for retrieval of the corresponding bibliographic information.

Description, Acquisition, Retrieval, and Correlation File. This is the only other public substructure search system, apart from CAS Online, that provides full access to the CAS Chemical Registry File. The DARC file, commercially available on-line from Telesystems-Questel, offered the first public on-line implementation of substructural searching of the CAS Chemical Registry System. The advantages and disadvantages of the CAS Online and DARC systems have been discussed (49).

Structure and Nomenclature Search System. This system links the collection of chemical databases found in the Chemical Information System (CIS), one of the first interactive systems for structure and substructure searching. References from the separate files can be retrieved by SANSS using CAS Registry Numbers, and the database of structures may be searched for structures or substructures. An adaptation of the SANSS software for substructure searching has been incorporated in the Drug Information System of the National Cancer Institute for its own use (54).

Numeric. Researchers routinely use reported numeric measurements and data in their work. Handbooks have been the primary source for locating this type of information, but numeric databases are now increasing in availability. Advantages of searching numeric databases on-line include ease of use, direct access to desired data, and ability to manipulate the information in the answer set.

Beilstein Handbook of Organic Chemistry. This reference (55) is one of the most significant collections of data in organic chemistry. The physical and chemical properties of organic compounds are tabulated in more than 500 fields. Most of these fields are searchable, and a sample of the record for chlorobenzene [108-90-7] is shown in Table 3.

Table 3. Sample Fields for Chlorobenzene in Beilstein

Name	Code	Number of references	Name	Code	Number of references
adsorption	CTADSM	14	infrared spectrum	IRS	43
association	CTASSM	47	ionization potential	IP	28
autonom name	AUN	1	kinematic viscosity	KV	5
azetrope	AZE	47	Lawson number	LN	1
Beilstein citation	SO	7	linear expansion coefficient	LEC	6
Beilstein preferred RN	BPR	1	liquid-liquid systems	CTLISM	29
boiling point	BP	103	liquid-solid systems	CTLSSM	14
bond moment	BM	2	liquid-vapor systems	CTLVSM	74
boundary surface phenomena	CTBSPM	19	liquid phase	CTLIQ	4
calorific data	CTCAL	4	liquid transition point	LTP	1
CAS Registry Number	RN	1	magnetic susceptibility	MSUS	12
chemical derivative	CDER	38	mass spectrum	CTMS	38
chemical name	CN	2	mechanical properties	CTMEC	15
chemical reaction	REA	1208	melting point	MP	24
circular dichroism	CDIC	1	molar polarization	MPOL	2
conformation	CTCFM	1	molar volume	MVOL	16
coupling phenomena	CTCPL	2	molecular energy	CTMEN	3
critical density	CRD	1	molecular formula	MF	1
critical pressure	CRP	4	molecular rotational constant	MRC	3
critical temperature	CRT	6	moment of inertia	MI	3
critical volume	CRV	3	nmr absorption	NMRA	69
cross-file reference	XREF	15	nmr data	CTNMR	6
crystal lattice parameter	CLP	4	nmr spectrum	NMRS	6
crystal phase	CTCRY	1	nuclear quadrupole coupling constant	NQC	4
crystal space group	CSG	4	nuclear quadrupole resonance	CTNQR	1
crystal system	CSYS	3	optical anisotropy	OA	8
crystal transition point	CTP	1	optical rotation dispersion	ORD	1
density (crystal)	DEN	1	optical rotatory power	ORP	1
density (liquid)	DEN	131	optics	CTOPT	15
dielectric constant	DIC	128	other source	OS	5
dielectric static constant	DISC	17	other spectroscopic methods	CTOSM	5
dipole moment	DM	111	polarographic half-wave potential	PHWP	11
dynamic viscosity	DV	26	preparation	PRE	127
electrical data	CTELE	17	purification	PUR	1
electrical polarizability	CTELP	2	Raman maximum	RAM	5
electrochemical behavior	CTECB	4	Raman spectrum	RAS	19
electronic absorption maximum	EAM	41	redox potential	RDXP	1
electronic absorption spectrum	EAS	40	refractive index	RI	129
electronic spectrum	CTESP	1	rotational spectrum	CTROT	4
emission spectrum	CTEMS	11	self-diffusion	SDIF	10
energy barrier of conformation	EBC	1	skeletal characteristics	CTSKC	4
energy of dissociation	EDIS	5	solubility	SLB	23
energy of MCS	CTENEM	44	solution behavior	CTSOLM	71
enthalpy of combustion	HCOM	3	stereo family	SF	1
enthalpy of formation	HFOR	5	structural data	CTGEN	1
enthalpy of fusion	HFUS	4	surface tension	ST	45
enthalpy of vaporization	HVAP	8	synonym	SY	1
entropy	SREF	3	thermal conductivity	TCND	7
ESR data	CTESR	3	transport phenomena	CTTRAM	38

formula weight	FW	1	unchecked data	CTUNCH	33
gas-phase behavior	CTGASM	16	use of compound	USC	2
heat capacity, C_p	CP	8	vapor pressure	VP	18
heat capacity, C_i	CV	1	vibrational spectrum	CTVIB	10
infrared maximum	IRM	20			

Gmelin Handbook of Inorganic and Organometallic Chemistry. This provides data similar to Beilstein for both inorganic and organometallic chemicals (56). A sample of the record for sodium metabisulfite [7681-57-4] is summarized in Table 4.

Table 4. Sample Fields for Sodium Metabisulfite in Gmelin

Name	Code	Number of references
CAS Registry Number	RN	3
chemical name	CN	1
component molecular formula	CMF	1
conformation and bonding models description of structure	CTCFM	1
crystal property description	CPD	2
enthalpy of formation	HFOR	1
formula weight	FW	1
general data	CTGEN	5
infrared spectrum	IRS	3
isotope search data	IFOR	1
linearized structure formula	LSF	1
molecular formula	MF	1
reaction	REA	48
solubility	SLB	58
spectroscopic information	CTSPE	4
uv and visible spectrum	UVS	2

Property Data Networks. These include the Materials Property Data Network, Inc. (MPD) (57) and Chemical Property Data Network (CPDN) and are available on STN. These networks provide menu access to numeric data on the performance of different materials and chemicals. Tables 5 and 6 summarize using the numeric files available on STN. NUMERIGUIDE is a data directory and property hierarchy support file produced by STN; it contains information on all properties available in the numeric files on STN.

Table 5. Materials Property Data Network

Database	Producer	Subject coverage
AAASD	The Aluminum Association, Inc.	mechanical and physical properties of commerical aluminum alloys
ALFRAC	U.S. Department of Commerce	testing procedures
ASMDATA	ASM International	specifications for composites, plastics, ferrous and nonferrous alloys, and metals
COPPERDATA	Copper Development Association, Inc.	numeric data for coppers and copper alloys
IPS	International Plastics Selector, D.A.T.A. Business Publishing	manufacturer-supplied data on commercial plastics
MARTUF	The National Materials Property Data Network, Inc.	toughness of steels
MDF	Materials Information ASM International	ferrous and nonferrous alloys
METALCREEP	The National Materials Property Data Network, Inc.	creep and rupture stress of aluminum alloys and steels
MH5	U.S. Department of Defense and Federal Aviation Administration	mechanical and physical properties for metallic aerospace materials
NISTCERAM	National Institute of Standards and Techology Gas Research Institute, Ceramics Division	mechanical, physical, electrical, thermal, corrosive, and oxidation properties for alumina nitride, beryllia, boron nitride, silicon carbide, silicon nitride, and zirconia
PDLCOM	William Andrew, Inc., Plastics Design Library	test data on the chemical compatibility and the environmental stress crack resistance of plastics
PLASPEC	D&S Data Resources	data on commercial plastics
STEELTUF	Electric Power Research Institute (EPRI) and PROD Materials Properties Council	toughness of more than 100 grades of steels used in the power industry
MPDSEARCH	The National Materials Property Data Network, Inc.	the <i>MPD Guide to Materials and Substances Data Sources</i> provides information about the materials property databases available on STN International
PLASNEWS	D & S Data Resources Plaspec	prices, market statistics, critical plastics industry news

Table 6. Chemical Property Data Network

Database	Producer	Subject coverage
DIPPR	American Institute of Chemical Engineers and Design Institute for Physical Property Data	numeric physical property data for commercially important chemicals and substances
HODOC	CRC Press, Inc.	numeric file representing the nine-volume 2nd ed. of the <i>CRC Handbook of Data on Organic Compounds</i>
HSDB	National Library of Medicine's Toxicology Information Program	toxicology and the environmental effects of chemicals
JANAF	U.S. Department of Commerce National Institute of Standards and Technology	chemical thermodynamic properties of inorganic substances and of organic substances containing only one or two carbon atoms

NISTFLUIDS	U.S. Department of Commerce National Institute of Standards and Technology	critically evaluated thermophysical and transport porperties for 12 important industrial fluids
NISTTHERMO	U.S. Department of Commerce National Institute of Standards and Technology	evaluated chemical thermodynamic properties of inorganic and organic substances containing one or two carbon atoms
POLYMAT	DKI, Deutsches Kunstoffinstitut and Fachinformationszentrum Chemie GmbH	data on commercially available plastic materials
RTECS	National Institute for Occupational Safety and Health (NIOSH)	registry of toxic effects of chemical substances; contains toxicity data and references commercially important substances
SPECINFO	Chemical Concepts GmbH	spectral data for a representative section of organic chemistry
TRCTHERMO	Thermodynamic Research Center, Texas A&M University, College Station, Texas	thermodynamic data

TDS NUMERICA. This is another source for numeric databases (58). This company provides different on-line databases and software for chemistry, engineering, and environmental data. A summary of its databases is contained in Table 7.

Table 7. TDS Numeric Databases

Database	Producer	Subject coverage
CHEMIST CHEMSAFE	Arthur D. Little, Inc. Physikalisch-Technische Bundesanstalt (PTB) (also available on STN)	predicts physical and environmental properties data for plant safety and accident prevention
DETERM Log P Database PPDS2	Dechema eV (also available on STN) Medicinal Chemistry Project at Pomona College National Engineering Laboratory (NEL, U.K.)	thermophysical properties and phase equilibrium data partition coefficients thermodynamic and phase equilibrium data, provides a modeling system based on several groups of physical properties data (six files)
TRC Vapor Pressure	Thermodynamic Research Center (also available on STN)	vapor pressure and boiling points

Other Databases. Available from different vendors (Table 8). For example, the researcher can obtain physical properties by using the *Merck Index Online* or the Dictionary of Organic Compounds available by Chapman and Hall Chemical Database. In DIALOG, numeric databases are collected under the name of CHEMPROP.

Table 8. Other Numeric Databases

Database	Producer	Vendor	Subject coverage
Chapman and Hall Chemical Database	Chapman and Hall, Ltd.	Dialog	<i>Dictionary of Organic Compounds</i> (5th ed.), <i>Dictionary of Organometallic Compounds</i> , <i>Carbohydrates</i> , <i>Amino Acids</i> , <i>Peptides</i> , <i>Dictionary of Antibiotics and Related Compounds</i> , and <i>Dictionary of Organophosphorus Compounds</i>
Computerized High Temperature Materials Properties Data-base EGIN-PLAST	Purdue University Guide de Choix Engin-Plast (France)	Purdue University (CINDAS) Mintel International Group Dialog	compiled from literature and evaluated reference data for high temperature materials mechanical, thermal, electrical, physical, processing, and flammability properties and applications for more than 4500 commercial plastics covers natural and synthetic polymeric materials
<i>Encyclopedia of Polymer Science and Engineering Online</i> Material Properties Bibliographic Data	John Wiley & Sons, Inc. Purdue University	Purdue University (CINDAS)	thermophysical, mechanical, and electronic properties of materials; bibliographic references and author index
Material Properties Numerical Data System	Purdue University	Purdue University (CINDAS)	evaluated data compiled, correlated, analyzed, and synthesized to generate values for the thermophysical, mechanical, and electrical properties of materials
MATUS	Engineering Information Co. (U.K.)	IPS, STN	mechanical, electrical, thermal processing properties
NAPRALERT (Natural Products ALERT)	University of Illinois at Chicago	STN	information on the pharmacology, biological activity, taxonomic distribution, medicine and chemistry of plant, microbial, and animal (including marine) extracts
NIST Update	High Tech Publishing Co.	NewsNet	news and information on activities of the U.S. National Institute of Standards and Technology (NIST)
NISTFLUIDS	NIST	STN	programs for calculating thermophysical and transport properties of cryogenic fluids
Plaspec Material Selection Database SPAO	Data Resources, Inc. Laboratoire National d'Essais (France)	Dialog, STN Teletel	detailed engineering and design data, chemical descriptions, and trade names for over 11,500 grades of plastics materials mechanical, electrical, thermal processing properties, and chemical resistance (Europe)
Agrochemical Handbook	The Royal Society of Chemistry, Cambridge (U.K.)	Dialog, Data-Star, Knowledge Index	information on the active components found in agrochemical products used worldwide, chemical names, including synonyms and trade names, CAS Registry number, molecular formula, molecular weight, manufacturers' names chemical and physical properties, toxicity, mode of action, activity, health, and safety
<i>The Merck Index Online</i>	Merck & Co., Inc.	Dialog, CIS, BRS,	the online counterpart to the printed 11th ed. of <i>The Merck Index</i> ; records contain molecular formulas and weights, systematic

TSCA	U.S. Environmental Protection Agency	Knowledge Index, Questel Dialog	chemical names, physical and toxicity data, therapeutic and commercial uses, and bibliographic citations information regarding chemical substances in commerce covered in the Toxic Substances Control Act
------	---	---------------------------------------	---

Cambridge. The Cambridge Structural Database is an integrated system of programs for searching, retrieving, and analyzing data on more than 96,000 organic and organometallic structures, which were determined by x-ray and neutron diffraction (59). About 15,000 compounds a year are being added to the database. The development of highly sophisticated x-ray diffraction equipment has contributed to greater precision in defining crystal structures and to more published crystallographic work. This database was developed by the Cambridge Crystallographic Centre of Cambridge, U.K. CAMBRIDGE has three basic files: classified bibliographic information, retrospective to 1935; evaluated numeric data for structures, since 1959; and chemical connectivity records, retrospective to 1935.

Because CAMBRIDGE is a compilation of all published crystallographic data, it is ideal for structure building and analysis. Text searching can be done in the bibliographic file. Substructure searching can be done using manual or graphic input. Once a structure or structure fragment has been found, related molecular and statistical data, such as bond lengths, bond angles, and coordinates, can be captured. The geometrical calculations and statistical analysis (GSTAT) feature can be used to generate further structural data, such as geometry of coordination spheres around specified atoms, calculation of centroids, vectors, and planes, user definition of atomic radii, fragment geometry and systematic tabulations, output of histograms, and dimensional scattergrams. CAMBRIDGE can also use PLUTO to generate a three-dimensional graphic output.

Examples of the versatility of CAMBRIDGE include its use in drug design to search for specific pharmacopores and the three-dimensional arrangement of chemical groups essential for biological activity (60). It has also been applied in molecular modeling (61) and amino acid and peptide studies. Development work is taking place to link CAMBRIDGE to MACCS-II to enhance the usefulness of both databases.

Spectra. The ability to consult collections of standard spectra is crucial in the analysis of unknown compounds. A long history of data collection efforts has been aimed at these applications. Among the best known of the published handbooks are the *Sadtler Spectral Data Sheets*, which include ir, Raman, and nmr spectra. An extensive bibliography of older hard-copy ir spectra is given in *The Coblenz Society Desk Book of Infrared Spectra* (62). Since the mid-1980s, comprehensive databases have been available in computerized form where the spectra themselves, not merely the bibliographic references, are searchable and displayable. The search algorithms vary considerably among the available systems; no algorithm standard exists (ca 1994), but several are under development (63,64). Expert systems, which assist in the automatic interpretation and identification of spectra, have existed for many years but are not commonly used (65). Computerized spectral databases are either local, PC-based, or public.

Local and PC-Based Databases. Local databases can be either free-standing or integrated with spectrometers. Many of the primary equipment manufacturers publish their own collections for use with their own instruments. Many free-standing local database systems are available for use on a PC or mainframe computer. Both types can have a database management system for building a user's personal database. Unless high standards are maintained in the accuracy of interpretation or calibration, the quality of these privately built databases can suffer. Also, disk space can limit the number of spectra stored. These disadvantages can be overcome by using on-line or CD-ROM databases, with their nearly high capacities and high quality; but these databases cannot currently be used to store and search personal data.

Nuclear Magnetic Resonance Spectroscopy. Bruker's database, designed for use with its spectrophotometers, contains 20,000 ¹³C-nmr and ¹H-nmr, as well as a combined nmr-ms database (66). Sadtler Laboratories markets a PC-based system that can search its collection of 30,000 ¹³C-nmr spectra by substructure as well as by peak assignments and by full spectrum (64). Other databases include one by Varian and a CD-ROM system containing polymer spectra produced by Tsukuba University, Japan. CSEARCH, a system developed at the University of Vienna by Robien, searches a database of almost 16,000 ¹³C-nmr. Molecular Design Limited (MDL) has adapted the Robien database to be searched in the MACCS and ISIS graphical display and search environment (63). Projects are under way to link the MDL system with the Sadtler library and its unique search capabilities.

A PC-based ¹H-nmr database, which includes full spectrum search capability, is being constructed by the Toyohashi University of Technology (67). SpecInfo, owned by Chemical Concepts, offers a 150,000 spectra library and database system for mainframe computers, which includes ¹H, ¹⁵N, ¹⁹F, ¹⁷O, ³¹P-nmr, and a large collection of ¹³C-nmr spectra compiled by Bremser at BASF (68,69). It also offers ¹¹B-nmr spectra compiled by Nuth at the University of Munich.

The National Chemical Laboratory for Industry (NCLI), Japan, has developed an integrated Spectral Database System (SDBS) which is available to users in Japan. All spectra were determined at NCLI under controlled conditions and are available on a PC CD-ROM or magnetic tape. The system has both ¹H-nmr (6000 compounds) and ¹³C-nmr spectra (5700 compounds), along with searching software. NCLI has also developed an integrated ¹³C-¹H-nmr system that can be used for two-dimensional data elucidation (70,71).

The Novosibirsk Institute of Organic Chemistry has developed a method for computer-aided retrieval of structural information from ¹H-nmr using its database of 50,000 spectra (72). Fraser Williams Ltd. (Scientific Systems) has special software to search its ¹⁹F-nmr database (73). Protein nmr data have been compiled into a relational database at the University of Wisconsin (74).

Infrared Spectroscopy. The Sadtler collection is the largest commercially available system with over 60,000 spectra, largely from prism and grating spectrometers. Fourier transform infrared (ftr) data are currently being added and will be searchable by substructure (63). Nicolet, in collaboration with Aldrich and Sigma, has developed separate databases, comprising over 10,000 ftr spectra each, of the compounds in the catalogues of the two companies. The EPA vapor-phase spectra collection is available through various ir instrument companies, and the SpecInfo system includes over 50,000 spectra from BASF and the Hummel ir Standards from Cologne University.

A computer file of about 19,000 peak wavenumbers and intensities, along with search software, is distributed by the Infrared Data Committee of Japan (IRDC). Donated spectra, which are evaluated by the Coblenz Society in collaboration with the Joint Committee on Atomic and Molecular Physical Data (JCAMP), are digitized and made available (64). Almost 25,000 ir spectra are available on the SDBS system developed by the NCLI as described. A project was initiated at the University of California, Riverside, in 1986 for the construction of a database of digitized ftr spectra. The team involved also developed algorithms for spectra evaluation (75). Other sources of spectral libraries include Sprouse Scientific, Aston Scientific, and the American Society for Testing and Materials (ASTM).

Mass Spectroscopy. A collection of 125,000 spectra is maintained at Cornell University and is available from John Wiley & Sons, Inc. (New York) on CD-ROM or magnetic tape. The spectra can be evaluated using a quality index algorithm (63,76). Software for use with the magnetic tape version to match unknowns is distributed by Cornell (77). The collection contains all available spectral information, including isotopically labeled derivatives, partial spectra, and multiple spectra of a single compound.

A second important database is the NIST EPA MSCD (formerly NBS EPA MSCD), jointly administered by the NIST, EPA, and the Mass Spectrometry Data Center (MSDC) of Nottingham, U.K. Many of the almost 55,000 spectra, especially those of environmental interest, are generated directly from the compounds at one of the sponsoring laboratories (67). A quality index algorithm, based on the one developed for the Wiley database, is available to evaluate the reliability of each spectra (64,78). Multiple spectra for a single compound are not included. Both the Wiley and NIST EPA MSDC databases incorporate older noncomputerized data collections as well as spectra from the literature (65). A merged version of these two databases is available from Wiley on CD-ROM or tape. SpecInfo offers a portion of the Wiley spectra as well as spectra from BASF, Max Plank Institute in Mulheim, ETH in Zurich, and the geo petrochemical collections from Kernforschungsanlage (KFA) Julich and Delft University. Fein-Marquart markets software for the PC called MASCOT, which includes a collection of about 1500 spectra of environmental compounds and forensic drugs (79). The NCLI integrated system, SDBS, has over 10,000 mass spectra (80,81).

Miscellaneous. NIST has a reference database of critically evaluated x-ray photoelectron and Auger spectral data, which is designed to run on PCs. It is searchable by spectral lines as well as by element, line energy, and chemical data (82). The Nuclear Quadrupole Resonance Spectra Database at Osaka University of over 10,000 records is available in an MS-DOS version (83). The NCLI system, SDBS, has esr and Raman spectra, along with nmr, ir, and ms data, as described.

Public On-Line Databases. These databases are accessible through on-line commercial services.

Chemical Information System. CIS was the first public on-line system to make a collection of spectra readily available. The initial work was done by the NIH, EPA, and other U.S. government agencies. Spectra data can be displayed in tabular or graphic format. Chemical substructures can be searched on all CIS databases using the Structure and Nomenclature Search System (SANSS). The structure query entry can be facilitated if the user has the PC front-end package, SuperStructure (71). A database of over 11,000 ¹³C-nmr spectra, collected by the Royal Dutch Chemical Society, can be searched interactively or by batch for shift, multiplicity, and intensity. The Infrared Search System (IRSS) allows retrieval spectra for known compounds as well as for unknown comparisons. IRSS contains over 4500 spectra for the EPA and the Boris Kidric Institute in the former Yugoslavia. CIS has two ways to search libraries of mass spectral data. The Mass Spectral Search System (MSSS) is used to search the NIST EPA MSDC database and a small collection of chemical ionization spectra. Besides chemical and formula information, MSSS can search by individual peaks or by full spectrum using a probability-based match. The Wiley Mass Spectral Search System (WMSSS) is similar to the MSSS but searches against the Wiley database.

SpecInfo/STN. A large proportion of the SpecInfo nmr, ir, ms databases described, along with additional data taken from the literature, are available through STN (Scientific and Technical Network International). Graphical access is possible through several PC emulation packages such as STN EXPRESS. MESSENGER command language is used along with several specialized commands, which were developed at BASF. CHESS is a chemical structure similarity search query. Coupling constants and nmr information can be searched using SPECAL, and spectra similarity searches are run with GETSPEC. Peak values can also be searched with all these techniques.

JICST/JOIS. The Japan Information Center for Science and Technology (JICST) Mass Spectral Database is accessible to users in Japan through the JICST Factual Database System (JOIS-F). The database uses the NIST EPA MSCD data collection supplemented by spectra from the Mass Spectrometry Society of Japan (84).

CAN/SND. The Canadian Scientific Numeric Database Service (CAN SND) is provided by the Canada Institute for Scientific and Technical Information (CISTI), a division of the National Research Council of Canada. It contains 140,000 ir spectra of 96,000 compounds. Entries consist of peak locations and some intensities. This system is searchable on-line using the SPIR (Search Program for Infrared Spectra) (85). Table 9 summarizes the available databases in the area of spectra.

Table 9. Summary of Spectra Databases

Producer	Type	Number of spectra	Availability	Notes
Aldrich-Nicolet	ftir	10,600	PC-based	compounds found in Aldrich catalogue
ASTM	ir	145,000		
Aston Scientific	ir			Quick-Search software
Boris Kidric Institute	ir	4,500	STN (vendor)	
Bruker	¹³ C-nmr	19,000	spectrometer-based	
	¹ H-nmr	900		
Canadian Institute for Scientific and Tech-nical Information	ir	140,000	CAN SND (vendor)	
Coblentz Society-JCAMP	ir	4,400	PC-based	
EPA	ir	3,300	STN (vendor) and spectro-photo meter	vapor phase
Fein-Marquart	ms	1,500	PC-based	environmental and forensic; MASCOT software
Fraser Williams (Scientific Systems Ltd.)	¹⁹ F-nmr		PC-based	PC-SABRE search software
Infrared Data Com-mittee of Japan	ir	19,000	PC-based	peak wavenumbers and intensities
Mass Spectroscopy Society of Japan	ms	6,000	JICST-JOIS	
NCLI, Japan	¹³ C-nmr	6,000 6,000 25,000	PC, CD-ROM, tape	integrated system available; Spectral Database System (SDBS) software
	¹ H-nmr ir ms	10,000		
	Raman esr			
NIST	xps		PC-based	
NIST EPA MSDC	ms	55,000	STN (vendor) PC-based, CD-ROM, tape	Wiley Quality Index algorithm
Novosibirsk Institute for Organic Chemistry	¹ H-nmr	50,000		
Osaka University	nqr	10,000	PC-based, CD-ROM, tape	
Royal Dutch Chemical Society	¹³ C-nmr	11,000	STN (vendor)	
Sadtler Laboratories	¹³ C-nmr, ir,	30,000 60,000	PC-based	PPC search software; substructure searchable; MACCS, ISIS link
	ftir			
Sigma-Nicolet	ftir	10,400	PC-based	compounds found in Sigma catalog
SpecInfo Chemical Concepts	¹³ C-nmr	100,000 13,000	STN (vendor)	Bremser-BASF collection Nuth, Munich University collection BASF; Hummel Standards BASF; portion of Wiley, Max Planck Inst.; ETH; geo petrochemical
	¹ H-nmr	1,000 900 1,900		
	¹⁵ N-nmr	2,200 9,000 50,000		
	¹⁷ O-nmr	160,000		
	¹⁹ F-nmr			
	³¹ P-nmr			
	¹¹ B-nmr ir ms			
Sprouse Scientific	ir			
Toyohashi University	¹ H-nmr	4,000	PC-based	full spectrum searchable
Tsukuba University	¹³ C-nmr		CD-ROM	polymers
University of California, Riverside	ftir			
University of Vienna	¹³ C-nmr	16,000	PPC-based	CSEARCH; MACCS, ISIS searchable
University of Wisconsin	¹ H-nmr			proteins
Varian	nmr		PC-based	
John Wiley & Sons, Cornell University	ms	125,000	CD-ROM, tape	isotopic labeled compound; Quality Index algorithm

Reactions. CASREACT File is a chemical reaction database containing over 118,500 records with reaction information derived from documents covered in the Organic Section of *Chemical Abstracts*. The file is available from STN. Coverage includes journals from 1985 to the present and patents from January 1991 to the present. The records contain structure diagrams for reactants and products, CAS Registry Numbers for all reactants, products, reagents, solvents, and catalysts, yields for many products, and textual reaction information. The reactants, reagents, and products are structure searchable with a single reaction query. Roles, reaction sites, and mapping of atoms between reactants and products are also structure searchable (86).

The database is updated biweekly with new records. The primary search terms in CASREACT are CAS Registry Numbers. Registry numbers may be input directly or may be contained in L-numbered answer sets and crossed over from the Registry file. Typically, all substances that are in CASREACT are flagged in the Registry file. In addition to the Registry numbers and structures to search the database, the feature Functional Groups allows use of different functional groups either as starting materials or products. An alphabetical listing of the CAS functional groups is as follows.

- Acetal
- acid halide
- acylmetal
- aldehyde
- alkene
- alkyne
- amide
- amidine
- anhydride
- azide
- azine
- aziriding
- azo
- benzenoid
- carbamate
- carbonate
- carboxylate
- carboxylic
- chloramine
- cyclic alc
- cyclic alkene
- cyclicdiene
- cyclicketone
- cyclopropyl
- diazo
- diene
- dihalide
- diimide
- disulfide
- dithioacetal
- dithiocarboxylate
- dithiocarboxylic
- enamine
- enol
- enol ether
- epoxide
- ether
- gem-dihalide
- glycol
- guanidine
- halide
- haloformate
- halohydrin
- hemiacetal
- heterocycle
- hydrazide
- hydrazine
- hydrazone
- hydroperoxide
- hydroxylamine
- imide
- imine
- iminoether
- isocyanate
- isonitrile
- ketal
- ketone
- lactam
- lactone
- metal halide
- metal hydride
- metal metal single
- metal phosphine
- metal sulfide
- metalarene
- metalcarbonyl
- metalcyclopentadienyl
- metallocarbocycle
- metalnitrogen
- metalnitosyl
- μ-carbonyl
- nitrile
- nitro

nitro
nitro
nitro
organometal
orthoester
oxime
peroxide
peroxyacid
phenol
phosphate
phosphite
phosphonate
phosphonium
phosphorus ylide
 π -alkene
 π -alkyne
 π -allyl
primary alc
primary amine
quaternary ammonium
quinone
secondary alc
secondary amine
selenide
selenol
silyl
silyl enol ether
sulfenyl halide
sulfide
sulfinic
sulfinic
sulfinyl halide
sulfonamide
sulfonate
sulfone
sulfonic
sulfonyl halide
sulfoxide
tertiary alc
tertiary amine
thioamide
thiocarboxylate1
thiocarboxylate2
thiol
thiolactam
thiophenol
triazene
trihalide
unsatdacid
unsatdaldehyde
unsatdamide
unsatdester
unsatdketone
unsatdnitrile
urea
and vinylhalide.

A sample of reduction of carboxylic acids to the corresponding primary alcohol is demonstrated in Figure 1.

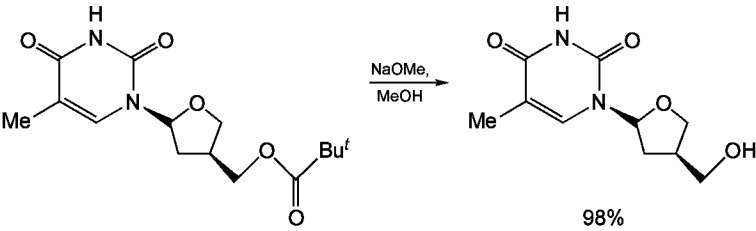


Fig. 1. Sample reduction of an ester group to the corresponding primary alcohol (CASREACT) (87): s(carboxylic or carboxylate) fg.rct (s) (primary alc) fg.pro.

CHEMINFORMRX from FIZ-Chemie Berlin and CHEMREACT from Springer-Verlag Infochem GmbH are other databases covering reaction information area. Reaction information is also available from Beilstein, Gmelin, and CA.

REACCS (reaction access system) is a specialized database management system for chemical reaction information. It is designed to store, search, retrieve, and display molecules, reactions, and the data associated with them (88). REACCS allows a variety of databases to be searched either individually or globally. The databases contain information such as bibliographic references, reaction conditions, yields, reagents, catalysts, and reaction class. Input may be either graphic or alphanumeric. In a REACCS database, a molecule is represented by a Stereochemical Extension of Mongan Algorithm (SEMA) name, two-dimensional coordinates, and structure keys. The SEMA name uniquely defines a molecule in terms of its atoms, their connectivity, stereochemistry,

charge, and isotopic properties (89).

REACCS is organized into six operational modes: MAIN and BUILD, which are used to draw molecular structures, build reactions, and construct graphic queries; and SEARCH, VIEWLIST, PLOT, and FORMS, which are used to create custom forms to display data associated with reactions and molecules. Each of the modes provides a characteristic menu and a set of options, which normally perform tasks that relate to the general function of that mode.

With its flexible and logical search language, REACCS can retrieve molecular structures, the atoms and bonds that are transformed in a reaction, relative and absolute stereochemistry, the role (reactant, product, solvent, or catalyst) of a molecule in a reaction, reaction data (eg, temperature and yield), literature references, and keyword descriptions of reaction types.

Data storage in REACCS is hierarchical: related data are stored separately but also are grouped under a single descriptive category. For example, in the Theilheimer database, the treename is the complete hierarchical name of a piece of data and is composed of three components: entity, parent datatypes or category of data, and field datatype. All REACCS databases include VARIATION. VARIATION is usually the highest parent datatype in the reaction hierarchy. VARIATION can be used to store more than one complete set of reaction data with a reaction. To keep track of the data associated with different variations or multiple reactants and products in the same reaction, line numbers are appended to some of the datatypes in a treename.

The four Build Menus in REACCS are Structure, Query, Top, and HighlightRxn. Structure menu contains the basic drawing commands used to construct the backbone of the structure. Query menu contains the commands used to add flexible structural parameters to the query. Top menu contains commands used to build reactions and to store and retrieve reactions, molecules, and graphic queries. HighlightRxn menu contains commands that apply atom atom mapping and reaction centers to the current reaction. Atom atom mapping is used to identify the reaction centers and increase accuracy and efficiency by letting the searcher specify that a particular atom in a reactant must correspond to a particular atom in the product.

Integrated Systems. Until recently, each of the numerous databases and sources of information available to chemists and technologists had to be searched individually, and selected results either printed for file storage or downloaded to an in-house or private computer system for easy future access.

Molecular Design Limited (MDL) (90) has marketed an Integrated Scientific Information System (ISIS), which provides the capability to query multiple systems, including binary, text, propriety, and relational databases across global networks, thereby providing transparent desktop access to multiple autonomous data sources. ISIS links databases, networking protocols, and computer platforms, providing a consistent environment across personal computer formats, by the use of three interrelated components: ISIS Draw, a state-of-the-art chemical drawing package which works on multiple computing platforms; ISIS Base, a database management program which allows storage of data and provides customized information that can be constructed on the desktop and stored or printed; and ISIS Host, which operates the server and interacts with clients and databases.

Patent Information and Search Methods/Services

Patents are unique as primary source documents because of their stylized format, specialized language, presence of legally significant claims, descriptive drawings, and frequent disclosure of chemical compositions as generic (Markush) structures. The term patent is commonly used to identify a broad family of patent publications at various stages of prosecution by national or regional issuing authorities, eg, unexamined applications, grants, and reissues. A comprehensive review of electronic databases that contain patent information, including legal aspects, is available (91) (see also PATENTS AND TRADE SECRETS).

Bibliographic Databases. A number of bibliographic and textual patent database producers have enhanced indexing to facilitate accurate retrieval with the use of controlled vocabulary, specific compound and substructure search capability, and additional explanatory text. Some database producers provide access via full text or inclusion of only some or all bibliographic details, title, and abstract, and others provide patent text or databases on CD-ROM for PC or network use. Newer resources allow users to input full or partial chemical structures graphically; the structures are automatically converted to appropriate functional group or topological coding for subsequent searching.

World Patents Index. WPI, produced by Derwent Publications, Ltd., contains records of patent publications from 32 issuing authorities around the world. Since 1970, all chemical technologies are included. Prior to then, content varies as follows: polymers from 1966, agricultural chemicals from 1965, and pharmaceuticals from 1963. Records contain bibliographic data and an abstract, which describes the novel features of the patent. The patent titles and abstracts in WPI are created by Derwent and generally give a more definitive description of the patent's content than the title and abstract, which appear on the document itself. Records also list equivalent or family member patents that have been identified as covering the same or related invention(s) (92).

WPI can be searched by bibliographic data, such as inventor, assignee (including standardized assignee codes), patent or application numbers or dates, and International Patent Classification, or by the in-depth indexing unique to the database. This indexing includes Derwent classification codes based on chemical composition, utility, or processes; chemical fragmentation codes (93), which define compounds by atoms, functional groups, ring systems, and carbon chains present; and polymer codes, which represent polymer type, monomers, processing, properties, and utility. About 2100 specific chemical compounds, which are important to the chemical and pharmaceutical industries, are searchable by Derwent Registry Numbers.

U.S. Patents. This file, produced by Derwent, Inc., covers U.S. patents from 1971 to the present. The database includes all bibliographic and front page information and the text of all claims. (From 1971 to 1974 the claims from many patents were not available from the United States Patent and Trademark Office (USPTO) source tapes, and therefore are not included.) The complete claim text can be searched from 1971 but can be printed only from 1982. Titles and patentee names are present in their original form, neither expanded nor standardized. There is no enhanced indexing. Examiner citations are directly searchable, and USPTO classification is updated when the tapes are received from the Patent Office.

CLAIMS. The CLAIMS databases are produced by IFI Plenum Data Corp. They cover only U.S. patents and include both bibliographic and auxiliary files.

CLAIMS BIBLIO includes an abstract and claim in addition to basic bibliographic information for chemical and chemically related U.S. patents from 1950 and for all patents from 1963. All claims are searchable and printable from 1971; claims for many patents are not available from 1971 to 1974. From 1972, many titles have been enhanced with additional keywords to describe the invention more clearly and to indicate the presence of a drawing; chemical structures have been converted so that they display in linear format. Many company names have been standardized, and USPTO classification is updated annually to reflect reclassification projects.

CLAIMS UNITERM adds enhanced indexing to the chemical and chemically related patent records; general terms to describe processes, properties, end products, etc; specific compound terms (over 15,000); and chemical fragment terms to describe generic compounds.

CLAIMS COMPREHENSIVE (subscriber access only) further enhances the UNITERM indexing from 1964 with roles for all indexed compounds and polymer class terms and with links and negation codes for the fragment terms describing chemical compounds (94).

CLAIMS CITATION includes examiner citations (prior references cited during prosecution of the patent application) against all U.S. patents from 1947. Citations are not directly searchable in the three bibliographic files.

CLAIMS REASSIGNMENT AND REEXAMINATION gives information on post-issue actions: reissues from 1950, reassignments from 1980, reexaminations from 1981, extensions from 1986, expiration for nonpayment of fees from 1985, and reinstatements.

CLAIMS CLASS contains the titles of the classes and subclasses of the *USPTO Manual of Classification*. It can be searched by title words to locate pertinent classes to use in the bibliographic files and by class subclass numbers to identify the classification assigned to a known patent. It is updated annually.

APIPAT. This is the patent database produced by the American Petroleum Institute and covers patents from 1964 of interest to the petrochemical industry, including petroleum refining, pollution control, uses of petrochemicals, and catalysts. Enhanced indexing includes terms applied from a hierarchical thesaurus with automatic posting to the broader terms in the hierarchy. Fragments called chemical aspects are linked to describe each compound, and the compounds are further linked to roles (eg, reactant or product) and use (eg, antioxidant or lubricant). ORBIT provides access to a merged APIPAT WPI file, which allows searchers to draw on the strengths of both databases without the need to search them separately (95).

EPAT. The European Patents Register is produced by the European Patent Office (EPO) and the Institut National de la Propriete Industrielle (INPI). The database provides bibliographic, including the first claim of granted (BI) patents, and legal status information on all European patents and published applications. Coverage is from June 1978, the beginning of EP publication, and now includes over 450,000 records. Titles on all records and

claims of granted patents are in French, English, and German. Abstracts, which have been included from 1988, are in the original language of the patent publication. The database is updated weekly on the day of publication.

JAPIO. This database is produced by the Japan Patent Information Organization and is based on the Patent Abstracts of Japan provided by the Japanese Patent Office. The database is updated monthly and contains all Kokai Tokyo Koho (published unexamined patent applications) published as of October 1976. Records appear in JAPIO approximately six months after publication of the unexamined patent application. English language abstracts are provided for the majority of applications filed by Japanese applicants. Applications by non-Japanese applications do not have abstracts, but bibliographic information is included. Searchable fields include the International Patent Office Classification and JAPIO classification (96).

INPADOC. This database has been produced by the European Patent Office since January 1991. It presently provides patent family and legal status information for over 16 million patent publications. Legal status, including oppositions, lapses or expiration for nonpayment of fees, and patent grants are provided for 11 issuing authorities. Patent family and bibliographic information is available for 56 national and regional patent offices, many providing coverage from as early as 1968. Titles appear in their original language and abstracts are not available. The database is updated weekly.

PHARM. This file contains bibliographic records of patents in the fields of pharmaceutical chemistry and biology. Coverage includes European, French, and U.S. patents from 1986, German and British from the 36th week of 1992, PCT from 1993, and French Special Patents for Medical Compositions (BMS) from 1961. Records emphasize the pharmaceutical aspects of the invention. PHARM is produced by INPI (97).

Full-Text Databases. The bibliographic databases discussed contain only a portion of the total information in the patent documents. Although several patent issuing authorities have, or are developing, full-text databases (98), remote on-line access to such databases has been limited. Two vendor-provided full-text patent databases are LEXPAT, produced by Mead Data Central, and PATFULL, produced by Dialog Information Services. The textual information for both of these is obtained from tapes available through the USPTO. These databases contain the full text of U.S. patents issued from 1974, but graphics, drawings, and chemical structures are not included. Searching can be based on words from the full document, USPTO Classifications, or International Patent Classifications. In 1994, Dialog announced the European Patents Fulltext database which contains the complete text of European published applications and patents, and the European PCT published applications. The database is produced by the European Patent Office and the text is in the original language of the application.

Chemical Substructure Databases. Several patent databases are searchable by chemical substructure (99). These are designed to give higher relevance of retrieval when searching chemical compounds than the bibliographic or full-text databases.

MPHARM is a companion database to the bibliographic PHARM. It contains the specific and generic structure records for compounds disclosed in patents included in the bibliographic database. Compound numbers located in MPHARM can be searched in PHARM to retrieve the corresponding bibliographic records (100).

WPIM (World Patents Index Markush), produced by Derwent Publications, Ltd., contains the specific and generic structure records for compounds in the patents included in Derwent Sections B (Farmdoc), C (Agdoc), and E (Chemdoc) since 1987. Sources include patents from 29 industrialized countries as well as European and PCT patents and also items from Research Disclosure and International Technology Disclosures. The compound numbers of relevant references found in WPIM can be searched in Derwent's WPI database to retrieve the corresponding bibliographic information.

MARPAT, produced by Chemical Abstracts Service, contains the generic structure records for patent publications since 1988, which are included in the CA file. Sources include patents from 26 countries plus EPO and PCT publications. Bibliographic records for retrieved references can be directly accessed in this database (101).

CD-ROM Databases. Since about 1989, CD-ROM format bibliographic and image patent databases have become available as current awareness, reference, or image storage and retrieval tools. The databases are designed for stand-alone PC or local network use, and in some cases they may be alternatives to retrieving patent information on-line. Image files dramatically reduce storage requirements compared to paper or microfilm. Records on each disk are searchable by the bibliographic information from patent documents or from weekly official patent office gazettes or bulletins. The USPTO and vendors are investigating user interest in full image technology subsets, eg, genetic engineering, biotechnology, and acid rain, in the CD-ROM format. Derwent Publications Ltd. provides *Documentation Abstracts* from 1992 as CD-ROM, with full-text abstracts and bibliographic information, including figures or structures. JAPIO provides image and index data as well as full-text for unexamined Japanese applications and utility models from 1987 to the present. Table 10 summarizes U.S. patent information resources available on CD-ROM.

Table 10. CD-ROM Databases

Database	Producer	Vendor	Description
APS (Automated Patent Searching)	U.S. Patent and Trademark Office (USPTO)	MicroPatent	covers U.S. granted patents 1975–present; first page and exemplary claim; updated monthly within two weeks of final issue date each month; cumulated to one disk three years
CASSIS	USPTO	USPTO Office of Electronic Data Conversion and Dissemination	CASSIS, the Classification and Search Support Information System of the USPTO, comprises three subfiles: CASSIS BIB, bibliographic information for utility patents from 1969 and for others from 1977; CASSIS CLASS, USPTO classification by patent number of class subclass; CASSIS ASSIST, index to U.S. Manual of Classification: U.S. Manual of Classification, Class Definitions; IPC, U.S. Classification Concordance; Manual of Patent Examining Procedure; Attorneys Agents Roster, etc
FullText	USPTO	MicroPatent	covers U.S. granted patents 1991–present; full text including examples, tables, but excluding drawings, structures; updated monthly, not cumulated; 12 disks year
OG PLUS	USPTO	Research Publications	CD-ROM version of the USPTO Official Gazette; covers 1990–present; includes searchable subfiles: PATENTS ISSUED, images of O.G. pages searchable by bibliographic fields and first page abstract; PATENT STATUS File, track-ing post-issuance actions, eg, reexaminations, corrections; and LIT ALERT, containing records of patent suits filed by U.S. District Courts with the USPTO; updated monthly; six disks year
Patent-Images	USPTO	MicroPatent	covers U.S. granted patents 1990–present; backfile to 1975 available; uses the same Patsoft software as the ESPACE products
PatentView	USPTO	MicroPatent	covers U.S. granted patents from 1986–present; backfile to 1974 to be available in 1994; customized subsets, eg, by company or technology, may be ordered

The European Patent office ESPACE series of CD-ROM products are summarized in Table 11.

Table 11. ESPACE Series of CD-ROM Databases

Database	Producer	Description
ESPACE-EP	MicroPatent, Research Publications,	contains full text and images of EP applications (EP-A) or granted

	EPO, WIPO	(EP-B) as separate collections; disks are up-dated weekly (85–90 year); covers 1989–present
ESPACE-FIRST	Micropatent, Research Publications EPO, WIPO	contains scanned images of first-page information of all EPO and PCT inter-national applications; bibliographic fields and patent titles are searchable in English, French, and German; disks are updated bimonthly (five disks year); covers 1989–present
ESPACE-UK	MicroPatent, Research Publications, EPO	contains full images of U.K. A documents; updated monthly; covers 1991–present
ESPACE-WORLD	MicroPatent, Research Publications, EPO	contains full text and images of PCT applications; titles are searchable in English, French, or German; updated bimonthly (35–40 disks year); covers 1981–present
ESPACE-ACCESS	MicroPatent, EPO	bibliographic data with abstracts for European (EPO) patent applications; disks are updated quarterly and are cumulative; covers 1978–present

MARKUSH TOPFRAG. Searching chemical compounds indexed in Derwent's World Patent Index database using fragmentation codes is a complex operation. It requires selecting appropriate codes from the coding manual or coding sheet and linking the codes in time- and subject-dependent search statements. In 1987 Derwent began to index chemical compounds based on an algorithm that allows for a topological search in file WPIM.

Derwent's TOPFRAG family of products is PC-based software that automates the selection of search codes and strategies. A chemical structure is input graphically using a drawing program, and the software generates the appropri-ate codes that define the structure. MARKUSH TOPFRAG, introduced in 1993, is the third generation of TOPFRAG and is the first designed to run under the WINDOWS environment. This software generates the chemical fragmentation codes for searching in WPI as well as the atoms, bonds, connections, etc, to be used in searching WPIM. The codes are generated in a line-by-line search strategy format. This strategy can be entered manually or, with some word processing, can be formatted to be uploaded via a communications software package.

Environmental and Safety Information and Search Methods and Services

Since the 1970s environmental and safety information and awareness have been characterized by legislation and regulation at both the state and federal levels. These actions have spurred a need to collect, organize, and retrieve information to aid in compliance with these laws. In an attempt to find a balance between public health, ecological balance and amenities, and industrial development, information supporting government, industry, and public actions has grown rapidly. The need to find information quickly has increased the demand for developing and using environmental databases.

Numerous reviews of environmental, safety, and health information sources have been published since 1981. A comprehensive review entitled "Environmental Information" was published in the *Annual Review of Information Science and Technology* (ARIST) in 1986 (102) and was updated in 1992. A three-part series entitled "Environment Online: the Greening of Databases" was published in *Database* magazine in August 1991, October 1991, and August 1992 (99,101,102). Part 1 covers general interest databases, Part 2, scientific and technical databases, and Part 3, business and regulatory databases; "Environment Online: Update 1993" appeared in the December 1993 issue of *Database* (103). This issue also reported a future trend for accessing electronically stored data for business purposes, eg, the use of the Internet telecommunications network to contact sources of information within the government directly (104). The article provides Internet connection numbers to such information sources as the Environmental Protection Agency (EPA), Online Library System, and the Department of Agriculture's National Agricultural Library, as well as toll-free numbers to EPA- and Department of Labor-sponsored bulletin boards. The government is considering providing a public on-line information system containing information sources and services within the government and indicating where to obtain them (ca 1994).

Comprehensive reviews of medical databases (105) and health and toxicological information systems (106), including search aids in each field, appeared in ARIST publications in 1983 and 1990. Toxicology information was reviewed in 1983 (103) and medical and health information in 1990 (100). Reviews of electronic government information (107) and engineering information systems (108) have also been published and provide an expansion of database knowledge for readers who require crossover information in these fields.

There are public and private databases. Public databases are produced by the government and private enterprise and are commercially available through database vendors, such as STN, DIALOG, BRS, ORBIT, and NLM, and through various universities. Private databases are produced by government agencies, corporations, or other organizations for in-house use by their employees or others affiliated with them. The information in these databases may be made available on a need-to-know basis to individuals or corporations in the public or private sectors. Knowledge of the existence of private databases is usually obtained by personal contact within an organization or thorough disclosure in published literature. Private industrial databases are not usually accessible by the public, although information contained in them on hazardous materials must be reported to the EPA under provisions of TSCA (Toxic Substances Control Act) Section 8e.

Public Databases. The most comprehensive list of publicly available databases is the two-volume *Gale Directory of Databases*, hereafter referred to as The Directory, published in 1993. Its index lists 140 databases under the subject heading Environment, 18 under Environmental Engineering, 17 under Environmental Health, 36 under Environmental Law, 50 under Waste Management, 61 under Toxicology, 39 under Industrial Hygiene, 63 under Hazardous Substances, 12 under Health Law, and 27 under Safety. Many entries are repeated from category to category, indicating information crossover within technologies and reinforcing the multidisciplinary nature of these technologies.

The environmental and safety databases are listed in Table 12. They are grouped by category, which is useful when searching the literature: Bibliographic Directory, Current Research Projects, Legal Regulatory (Table 13), Numeric Data, and Newsletters. The Directory may be referenced for details of specific subject entries and for database availability. Databases, such as Chemical Abstracts, Biological Abstracts, and Engineering Index, are not included; for their details, see The Directory.

Table 12. Environmental and Safety Databases

Database	Producer	Subject
	Bibliographic directory	
Acid Rain	Reed Reference Publishing Group	acid rain
	Bowker A&I Publishing	
ACIDOC	Quebec Ministere de l'Environnement	acid rain
AGRIS	United Nations Food and Agricultural Organization (FAO)	agriculture
Applied Science & Technology Index	H. W. Wilson Co.	environment
AQUALINE	WRc plc	water resources
AQUAREF	Environment Canada	water resources
BALTIC	Sweden Statens Naturvardsverk (SNV)	Baltic Sea
BIOLIS	Informationszentrum für Biologie	environmental biology
Biological & Agricultural Index	H. W. Wilson Co.	environmental science
BNA Books	Bureau of National Affairs (BNA)	employment law

Chemical Activity Status Report (CASR)	Chemical Information Systems, Inc. (BNA)	chemicals under EPA review
Coal Database	IEA Coal Research	coal science
Current Awareness in Biological Sciences	Pergamon Press Plc.	toxicology, ecology
Current Contents Search	Institute for Scientific Information (ISI)	environmental sciences
DECHEMA DETEQ	Dechema	German environmental equipment technology
Directory of Occupational Safety & Health (OSH)	Labour Canada	Canadian OSH-related legislation
Eastern European Energy Report	Strategic Marketing	Eastern European environment industries
ECOCERVED	Ecocerved	Italian environment
Ecology Abstracts	Cambridge Scientific Abstracts (CSA)	ecology, environment
EDF-DOC	Electricite de France (EDF)	electric power
ELIAS	Environment Canada	Canadian environment
EMBASE	Elsevier Science Publishers	biomedical, OSH, toxicology
ENERGIE	FIZ Karlsruhe	energy-related aspects of environmental and biomedical sciences
Energy Science & Tech-nology (EST)	U.S. Department of Energy (DOE)	energy conservation
Enviroline	Reed Reference Pub-lishing Group	natural resources
Environment Protection	VINITI (Vsesoyuznyi Insitiut Nauchnoy i Technicheskoy Infor-matsii)	environmental protection
Environmental Biblio-graphy	International Academy at Santa Barbara	environment
Environmental Mutagen Information Center Data Base	Oak Ridge National Laboratory	physical agents tested for mutagenic activity
Environmental Resources Technology	Petroleum Abstracts	petroleum exploration, production, transport
Environmental Teratology Information Center Database	Oak Ridge National Laboratory	teratology
Facilities Index System	U.S. Environmental Protection Agency (EPA)	EPA-regulated sites
GEOBASE	Elsevier Geo Abstracts	earth sciences
Hazardous Communi-cation Standard Compliance Manual Database	BNA	OSHA hazards
Health and Safety Science Abstracts	CSA	hazards control
ISTP&B Search	ISI	environmental science
MSDS FTSS	Canadian Centre for Occupational Health and Safety (CCOHS)	chemicals
National Environmental Data Referral Service Database (NEDRES)	U.S. National Environ-mental, Satellite, Data and Information Service (NESDIS)	environmental data
NATUR	SNV	Swedish environmental issues
Occupational Health & Safety: Seven Critical Issues	BNA	OSHA concerns
Oceanic Abstracts	CSA	marine sciences
OIL	Oljedirektoratet	Scandinavian oil industry
Pollution Abstracts	CSA	pollution
POWER	DOE	energy production
Report on Defense Plant Wastes	Business Publishers, Inc. (BPI)	waste management
REPROTOX	Columbia Hospital for Women	chemical effects on human reproduction
SIREN	Portugal Centro de Estudosem Economia Energia dos Trans-portese e Ambiente	Portuguese environment
Standards & Direct-ories Normes et Reportoires	Canadian Centre for Occupational Health	Canadian occupational health and safety
Superfund	Pasha Publications, Inc.	hazardous waste cleanup
Tanker	Institut Francais du Petrole (IFP)	shipping accidents with oil released
Toxicological Aspects of Environmental Health	Biosis	pollution effects on environment
Toxicology Abstracts	CSA	toxicology
TOXLINE	U.S. National Library of Medicine (NLM)	toxicology, pesticides, mutagens
UMEDIA	Institut der Deutschen Wirtschaft	environmental issues
Umwelt-Datenbank	Online Gesellschaft fir Informationsvermitt-lung mbH	German environmental protection products
Umweltliteraturdatenbank (ULIDAT)	Deutsches Umweltbundesamt	German environmental topics
VROMDOC	Netherlands Ministry of Housing	Dutch land registry
Waste Management & Resource Recovery	International Research and Evaluation (IRE)	waste management
WATERNET	American Water Works Association	wastewater treatment
Coal Research Projects Data Base	Current research projects	
Energy Research in Progress (ERIP)	IEA Coal Research	coal technology
Environmental Research Projects (ENREP)	Energy Research Development Corp.	Australian energy and conservation
EUREKA	Commission of the European Com-munities	European Community (EC) environment
Umweltforschungs-datenbank (UFORDAT)	Eureka Secretariat	EC's Eureka program
	Deutsches Umweltbun-desamt	Austrian environment
	Numeric data	
AQUA II	Infochem Computer Services Ltd	thermodynamic properties of aqueous solutions
BAKER	J. T. Baker, Inc.	material safety data sheets (MSDS) for 1500 chemicals
CERCLIS Database of Hazardous Waste Sites	EPA and CIS	hazardous substances releases reported to the EPA
CHEMEST	Technical Database Services, Inc. (TDS)	properties of pharmaceuticals and chemicals
Chemical Evaluation Search & Retrieval System (CESARS)	Michigan State Depart-ment of Natural Resources	toxicological data on 370 toxic chemicals

Chemical Hazards Re-sponse Information System (CHRIS)	U.S. Coast Guard	water transport of hazardous chemicals
Chemical Identification of Medicine File (ChemID)	NLM	nomenclature and structure of 200,000 chemicals in NLM file
Environmental Chemicals Data and Information Network (ECDIN)	Commission of the Euro-pean Communities	toxicity of 122,400 chemicals in the environment
Environmental Fate (ENVIROFATE)	EPA	fate of 800 chemicals released to the environment
Environmental Fate Databases	Syracuse Research Corp.	DATALOG, CHEMFATE, BIOLOG, BIODEC files on fate of organic chemicals released into the environment
Hazardline	Occupational Health Services (OHS)	regulatory data on 90,000 hazardous chemicals
Hazardous Chemicals Information and Disposal Database (HAZINF)	University of Alberta	handling instructions for hazardous substances
National Analysis of Trends in Emergencies System (NATES)	Environment Canada	hazardous spill incidents in Canada
OHS Material Safety Data Sheets (OHS MSDS)	OHS	information on 85,000 OSHA-documented chemicals
OHS MSDS Summary Sheet	OHS	information on 10,000 chemicals
Oil and Hazardous Materials Technical Assistance Data System (OHM-TADS)	EPA	technical support for dealing with dangers from oil or hazardous substances
Radioactivity Environ-mental Monitoring (REM)	EC	radioactivity data from the EC relating to food chain contamination
Toxic Chemical Release Inventory (TRI)	EPA	estimated releases of toxic chemicals from 20,000 industrial sites
TSCA Plant and Production Data (TSCAPP)	EPA	production data for TSCA Chemicals
TSCA Test Submissions (TSCATS)	EPA	4200 TSCA chemicals
	Newsletters	
BNA Chemical Regu-lation Daily	BNA	legislation on pesticides, chemicals
BNA Daily News	BNA	U.S. government legal issues
BNA Environmental Law Update	BNA	legal actions on EPA rules, Super-fund cleanup, wetlands, etc
BNA International Environment Reporter	BNA	global pollution control legis-lation, conferences, treaties
BNA Occupational Safety & Health Daily	BNA	legal issues affecting occupational health and safety
Brazil Watch	Orbis Publications, Inc.	Brazilian politics, economics, business
Business and the Environment	Cutter Information Corp.	global environmental policies
California Planning & Development Report	Torff Fulton Associates	California land-use regulations and growth control
Environment Watch: Latin America	Cutter Information Corp.	Latin American environmental initiatives
Environment Week	King Communication Group, Inc.	environmental issues, acid rain, recycling, greenhouse effect
Environmental Business Journal	Environmental Business Publishing, Inc.	environmental industry business
Global Environmental Change Report	Cutter Information Corp.	global warming, ozone depletion, acid rain, deforestation
Golob's Oil Pollution Bulletin	World Information Systems	global oil spills: control, prevention, pollution
Greenhouse Effect Report	Business Publishers, Inc. (BPI)	global climatic warming
Ground Water Monitor	BPI	groundwater legal issues, hazardous waste disposal
Industrial Environment	Worldwide Videotex	industrial environment improvement measures
Industrial Health & Hazards Update (IH&HU)	Merton Allen Associates	industrial health and hazards regulations
Louisiana Industry Environmental Alert	Environmental Compliance Reporter, Inc.	industrial regulations and actions in Louisiana
New Jersey Industry Environmental Alert	Environmental Compliance Reporter, Inc.	DEP and EPA policies and effects on New Jersey environmental issues
Nuclear Waste News	BPI	nuclear waste management, safety, security
Occupational Health & Safety Letter	BPI	health and safety in the workplace
Occupational Safety & Health Reporter	BNA	worker safety and health issues
Ozone Depletion Network Online TODAY	Environmental Infor-mation Networks	stratospheric ozone depletion
PressNet Environmental Reports	PressNet Systems, Inc.	state and local environmental issues
Texas Industry Environmental Alert	Environmental Compliance Reporter, Inc.	regulations and EPA actions affecting Texas industries
Toxic Materials News	BPI	legislation related to TSCA
Toxics News	Capitol Reports	hazardous wastes, air and water quality
U.N. Conference on Environment and Development	United Nations Conference on Environment and Development	press releases, documents, speeches from this conference
Waste Information Digests	International Academy at Santa Barbara	waste management, recycling
Waste Treatment Technology News	Business Communications Company (BCC)	handling and management of hazardous waste
World Environment Outlook	BPI	global environmental issues

Table 13. Legal/Regulatory Environmental Databases

Producer	Subject	Coverage
BNA California Environment Daily	Bureau of National Affairs (BNA)	California environmental law
BNA Chemical Regulation Daily	BNA	pesticides, chemicals, biotechnology
BNA Chemical Regulation Reporter	BNA	chemicals, pesticides, hazardous wastes
BNA Daily Environment Report	BNA	state international environmental issues
BNA Environmental Law Database	BNA	chemistry, pesticides, environment
BNA Environmental Law Update	BNA	environmental policy
BNA International Environment Daily	BNA	pollution, waste management
CELDS Environmental Regulations	University of Illinois at	U.S. environmental regulations

	Champagne-Urbana	
Daily Report for Executives	BNA	government regulations
Environment Reporter	BNA	pollution, hazardous waste, environment
Environmental Compliance Update	High Tech Publishing Co.	business compliance with environmental standards
Environmental Health News (EHN)	Occupational Health Services (OHS)	environmental and occupational health
Environmental Law Reporter	Environmental Law Institute	U.S. environmental law
Guide to Federal Environmental Laws	BNA	federal environmental laws for human resource professionals
Statens Naturvardsverk	DAFA Data AB	Swedish National Environmental Protection Board statutes
Toxic Law Reporter	BNA	hazardous waste laws and news
WESTLAW Environmental	West Publishing Co.	state environmental regulations in U.S.
Administrative Law Database		
WESTLAW Environmental Law Library	West Publishing Co.	U.S. environmental law
WESTLAW Topical Highlights	West Publishing Co.	federal and state court decisions on environmental law

Private (EPA) Databases. The U.S. EPA maintains a list of approximately 600 current information systems, as well as some of the models and databases used within the organization. The list is published in *Information Systems Inventory* (ISI) which is updated yearly and maintained by the Information Management and Services Division of the Office of Information Resources Management (109). ISI lists the system name and acronym, system level, responsible organization, contact person, legislative authorities, database descriptors, access information, hardware and software, system abstract, and keywords.

ISI is available in hard copy and electronically at EPA's headquarters and regional libraries, and through the National Technical Information Service (NTIS). The electronic form may be installed on IBM PC-compatible computers or placed on local area networks, and run under Microsoft WINDOWS or WordPerfect's Library program. The Macintosh version is no longer available. The 1993 update will include the ISI hardcopy, PC disks, and the PC system user manual. EPA also publishes ACCESS EPA, which provides sources of information, databases, and publications within the EPA. Chapter 5 of that publication includes important environmental databases in air and solid waste, pesticides and toxic substances, water, and cross-program (110). EPA also provides databases accessible through EPA libraries, which describe the private EPA and commercial databases available to library users (111).

On-Line Search Aids

MACCS-II. The Molecular Access System is a chemical information management system from Molecular Design Limited (MDL), San Leandro, California. It offers menu-driven graphical input for building, maintaining, and accessing chemical structures and any associated data, eg, chemical and physical properties, biological activity, toxicity data, pricing, safety, and supplier information. Substructure searches are done by drawing the compound with a mouse or light pen and activating the search process. The software interprets the drawn atoms, bonds, and stereochemistry of a chemical structure; retrieves appropriate compounds; and graphically displays them with their corresponding data. MACCS-II allows for customization of the environment to suit various applications. Because it is designed for both mini- and main-frame computers, this system is suited for developing proprietary personal or corporate databases in which unique or new compounds can be registered together with the data of interest (112). Several companies such as Sandoz and Zeneca have taken advantage of this feature.

The following chemical databases are available for searching in MACCS-II. *Chemical Directory Database* contains a combined catalogue of 66 commercial suppliers of more than 77,000 organic chemicals. *MACCS-II Drug Data Report*, based on the Prous Drug Data Report, includes 39,000 compounds with information on therapeutic indication, biological action, phase of development, related patents, and literature references. *MUSE Database*, the tutorial database for MACCS-II, contains over 100 compounds.

MDL has also developed a special version of MACCS-II called MACCS3D. This is used for storing and searching three-dimensional chemical structures and related data. MACCS3D provides convenient graphical input for building three-dimensional queries based on atom distances, bond angles, torsions, planes, centroids, and excluded volumes. Special facilities are also provided for viewing three-dimensional images of retrieved compounds and related data.

The MACCS3D databases available for searching are *MUSE3D Database* which contains over 700 compounds and is useful in learning to use MACCS3D; *FCD3D* is the Fine Chemicals Directory Database of commercial chemicals; and *CMC3D* is the Comprehensive Medical Chemistry Drug Compendium containing about 6000 compounds. *MDDR3D* is the Drug Data Report, Volume 1, which includes biological data.

MACCS-II enables direct interface with other database management systems, such as the Relational Database Management System (RDBMS) and Oracle, so that databases which contain text and numeric data for which special interfaces are normally needed can be constructed. For example, an Oracle MACCS-II linked system is currently being used by the National Institute on Drug Abuse (113) to develop a database that will allow scientists to determine the molecular structures of cocaine and other controlled substances as well as designer drugs.

Optical Disk-Based Information and Document Image Systems

Optical-based storage technology has joined paper, microfilm, and electronic magnetic technologies as another medium for the storage, retrieval, and management of information (114–116) (see INFORMATION STORAGE MATERIALS). Optical media differ from magnetic media in that the information is encoded and read by means of laser optics. Information stored on optical disk may be either in a searchable text format (ASCII) or in a format containing only bit-mapped images, usually obtained as output from a scanner. Through the scanning and digitization process, pages that consist of printed text, graphics, photographs, drawings, handwriting, tables, etc, are converted to their binary representation and are stored as bit-mapped images on the optical media. The information stored on optical disk often has counterparts in other formats, such as printed publications or on-line databases or files. Advantages of document imaging systems for complementing, enhancing, or replacing traditional paper- or microfilm-based systems include increased storage capacity, ease of access via automated retrieval, simultaneous searching and viewing at multiple workstations, speed of access and delivery resulting in productivity gains, improved customer service, document security, document integrity via preservation and elimination of lost or misfiled documents, and networking and integration capabilities for these systems (114). Among the types of optical media that have been developed, the two most common for information storage and retrieval systems are both optical disk-based systems, namely CD-ROM and WORM (write once read many).

Searchable text information on optical disk can range from bibliographic databases, abstracts, and indexes, to full texts of books, journals, and reference publications, to numeric data. Printed text, graphics, drawings, photographs, spectra, and tabular information may also be stored on optical disk as images. Applications of document-image systems are numerous and span all types of industries and organizations. The potential for integrating, linking, and networking document image systems with other existing electronic databases and systems is being explored in efforts to bring the virtual library to reality.

Several projects such as CORE (Chemistry On-Line Retrieval Experiment) at Cornell University, Project Mercury at Carnegie-Mellon University, RightPages at AT&T, and Red Sage at the University of California, San Francisco are in progress and illustrate the issues arising from implementation of on-line information systems that combine text and image (6,116).

The objectives of these projects are to investigate how new desktop interfaces to scientific information affect the way scientists access and use the information from databases, journals, and other proprietary and published literature and to address the technical aspects of assembling these seamless networks and user-friendly interfaces. These and other challenges that lie ahead for users, publishers, and suppliers of CD-ROM, optical disk, and document image storage systems encompass some of the same issues that are affecting other areas of information science and technology. The issues of standardization, copyright, and the admissibility of proprietary electronic and optical records as legal documents are not being resolved as quickly as the underlying storage and retrieval technologies are advancing.

Optical media have the capacity to handle much larger amounts of digitized information than equivalent sizes of magnetic media. These capacities

are increasing each year. A standard 4.75-inch CD-ROM can hold more than 680 MB (megabytes) of data, roughly the equivalent of 250,000 pages of text. Storage capacities of WORM disks range from 250 MB for 5.25-inch disks to more than 14 GB (gigabytes) for 14-inch ones; this translates into 20,000 and 130,000 imaged pages, respectively. Whereas CD-ROM has emerged as a suitable medium for the information publishing industry, WORM applications are used mainly for archival document image storage systems.

Publications on CD-ROM. The *Gale Directory of Databases*, Volume 2, contains a comprehensive listing of available CD-ROM based products (117). *Analytical Abstracts*, *Chemical Abstracts 12th Collective Abstracts and Index*, *MEDLINE*, *Chemistry Citation Index*, the *CASurveyor* series (separate disks on specific areas of chemistry), and the *PolTax* series (pollution and toxicology) are among some of the bibliographic databases and indexes available in CD-ROM versions. Full-text sources on CD-ROM include catalogues: *Aldrichem Data Search*; encyclopedias: *Kirk-Othmer Encyclopedia of Chemical Technology* and *Polymer Encyclopedia* (complete text of the *Encyclopedia of Polymer Science and Technology*); dictionaries: *CHCD Dictionary of Organic Chemistry* and *CHCD Dictionary of Inorganic Chemistry*; and directories: *Chem Sources*. Examples of numeric data sources available on CD-ROM are toxic and hazardous chemicals information in *Chem-Bank*, x-ray powder diffraction patterns in *Powder Diffraction File*, and properties of plastic materials in *CenBASE/Materials*. *Beilstein Current Facts in Chemistry* on CD-ROM allows searching of current bibliographic citations on organic chemistry by chemical structure and other physical and chemical properties. Some journals on CD-ROM are available in full-text searchable format; for example, *The Lancet* (1989–1994). The *ADONIS* collection of CD-ROMs contains full-text images of all articles, letters, and abstracts from more than 490 biomedical journals. Other commercial document image collections on CD-ROM include U.S. patents on *PatentImages* and *PatentView*, European patent applications on *ESPACE*, and Department of Defense Standards on *DoD Standardization Service*. The *Worldwide Standards Service* on CD-ROM contains a comprehensive index to standards and specifications from more than 375 international organizations and associations; the full standards as images on CD-ROM are also available as part of this service.

CORE. The CORE Electronic Chemistry Library is a joint project of Cornell University, OCLC (On-line Computer Library Center), Bell Communications Research (Bellcore), and the American Chemical Society. The CORE database will contain the full text of American Chemical Society Journals from 1980, associated information from Chemical Abstracts Service, and selected reference texts. It will provide machine-readable text that can be searched and displayed, graphical representations of equations and figures, and full-page document images. The project will examine the performance obtained by the use of a traditional printed index as compared with a hypertext system (SUPERBOOK) and a document retrieval system (Pixlook) (6,116).

Project Mercury. This system, at Carnegie-Mellon University, implements electronic library architecture through distributed rather than centralized computing. Elsevier and the Institute of Electronic and Electrical Engineers are also helping to distribute full-page images to users' desktop computers across the campus network. In this project a search is performed in the INSPEC database, citations are selected on the screen, and the full article appears for viewing. Detailed statistics are gathered to provide the publishers with information that helps them devise marketing, pricing, and delivery strategies (118).

RightPages. The RightPages image-based electronic library, developed for users at AT&T Bell Laboratories, is an on-line version of the periodical shelves of a conventional library. An added feature is that RightPages alerts users to the arrival of new journal articles matching a specified profile and enables them to examine pages in the article and browse other articles in the database (119). AT&T is also collaborating with Springer-Verlag in another electronic library project at the University of California, called Red Sage (6). Forty Springer-Verlag journals in the areas of molecular biology and radiology will be scanned into a bit-mapped image system and transformed into searchable text via OCR (optical character recognition). The user interface for this project will be AT&T's RightPages and will contain a hyper-paper feature that will allow users to view pages containing figures immediately by pointing to the citations for those figures.

Private Bibliographic and Text Databases

Personal computers have introduced new ways to handle private bibliographic and text files. The most important factors to consider to achieve satisfactory results in building a bibliographic or text database are the type of information to be stored and the needs of the user. Types of information include correspondence, research results and documentation, meeting notes, and bibliographic references. Needs of the user to be considered should include the potential number of users of the database, restrictions for the access and display of the information because of privacy or proprietary reasons, and the retrieval mechanisms (eg, by keyword, authority list, controlled vocabulary, author, title, date, or other document or information attributes). In addition, criteria for selecting and encoding information for the database need to be established.

The potential size of the database and the number of users will influence the decision to use free text, keywords, or natural language vs a thesaurus-controlled vocabulary for subject description or analysis. In free text or natural language, individuals may select keywords or they may be automatically generated by the software used. Controlled vocabulary assignment is done using an authority list or thesaurus, which is subsequently used for retrieval. For example, the controlled term "preparation" could be used to retrieve documents on synthesis, manufacturing, preparing, formation, or reactions of. In free-text searching, all of the above words would need to be searched. Controlled vocabulary assignment requires more effort to input but less effort to search; free-text keyword assignment requires less effort to input and more to search. Additional information is available on indexing or term assignment (120–125).

If the database is to contain published literature, many software packages have the capability to download or copy from on-line databases. Also, most terminal emulation software packages have a copy or capture feature. When downloading from commercial databases, license agreements and copyright requirements must be honored (check with the database vendors as to their specific licensing agreements). In addition to published literature, proprietary information can be added to the database by copying or downloading from an in-house computer system (125).

The type of hardware or computer system to be used and the potential size of the database should also be considered. For some databases, a personal computer may be adequate. For others, especially if there will be multiple users, a mainframe computer or a network of personal computers may be required. Storage capacity and response time are parameters that should be considered. However, computer technology changes so rapidly that vendors and computer experts should be consulted when building any personal databases.

Another factor frequently overlooked in private database creation is commitment to the support and maintenance of the database. Support involves training users, solving software and hardware problems, and upgrading the software when new features become available or are needed by the end user. Maintenance of the database includes adding new information, deleting information no longer wanted, and correcting information in the database when errors are detected. Once the needs and requirements are documented, available software products should be surveyed to determine which one is most suitable and whether customization to fit the user's needs is required. Customization may include defining the format of the field or area in the database in which the information is to be stored. Some examples are the format for the name of an author, eg, last name and initials, or last name and full first name; how many characters should be allowed in a field for a title or abstract; and the format for a date, eg, year month day or day month year. Depending on the software selected and the skills of the database user, customizing can be done by the user or someone trained in computer technology. Libraries, research groups, and individuals can use a variety of software products to develop private databases in their specific areas of interest (126,127). Software packages available for building databases range from generic personal computer database management systems, which require customizing to software designed specifically for bibliographic files. Most of the software is for Macintosh or IBM compatible PCs. Some examples are PROCITE (128), NOTEBOOK II (129), ENDDNOTE (130), LIBRARY MASTER (131), PERSONAL FILE SYSTEM (132), ASKSAM (133), BASISPLUS (134), and PERSONAL LIBRARIAN (135). A buyer's guide to software products for management of information by industrial scientists is available (128). Vendors and their product literature and product reviews in library, information science, and personal computing journals are good sources of information. Private databases offer a fast, convenient, and economic tool to aid researchers in planning, interpreting, and reporting their results.

BIBLIOGRAPHY

"Literature, Mechanized Searching" in *ECT* 1st ed., Vol. 8, pp. 449–467, by J. W. Perry, Bjorksten Research Laboratories, and R. S. Casey, W. A. Sheaffer Pen Co.; "Literature of Chemistry and Chemical Technology" in *ECT* 2nd ed., Vol. 12, pp. 500–529, by T. J. Devlin, Esso Production Research Co., and B. H. Weil, Esso Research and Engineering Co.; "Information Retrieval Services and Methods" in *ECT* 2nd ed., Suppl. Vol., pp. 510–535, by E. Garfield and

- C. E. Granito, Institute for Scientific Information, and A. E. Petrarca, Ohio State University, "Information Retrieval" in *ECT* 3rd ed., Vol. 13, pp. 278–336, by M. H. Graham and L. Y. Stroumtsos, Exxon Research and Engineering Co., A. B. Lamy, Essochem Europe, Inc., and B. Lawrence, Exxon Corp.
1. H. Skolnik, *The Literature Matrix of Chemistry*, John Wiley & Sons, Inc., New York, 1982, p. vi.
 2. G. Wiggins, *Chemical Information Sources*, McGraw-Hill Book Co., Inc., New York, 1991.
 3. R. Maizell, *How to Find Chemical Information*, John Wiley & Sons, Inc., New York, 1987.
 4. J. Branin, *Collection Building* **9**(3–4), 20 (1989).
 5. M. Khalil, *Library J.* **118**(2), 43 (1993).
 6. S. Borman, *C. & E. News* **71**(24), 10 (1993).
 7. *Advanced Technology Libraries* **20**(12), 3 (1991).
 8. *The High-Performance Computing Act of 1991*, PL 102–194, 105 STAT. 1594, Washington, D.C.
 9. R. Tennant, J. Ober, and A. G. Lipow, *Crossing the Internet Threshold: An Instructional Handbook*, Library Solutions Press, San Carlos, Calif., 1993.
 10. E. Krol, *The Whole Internet User's Guide & Catalog*, O'Reilly & Associates, Inc., Sebastopol, Calif., 1992.
 11. S. Fishman, *The Copyright Handbook*, Nolo Press, Berkeley, Calif., 1992.
 12. L. L. Schaper, *Proceedings of the 12th National Online Meeting, 1991*, Learned Information, Medford, N.J., 1991, pp. 349–352.
 13. J. Garrett, *Online* **15**(2), 22 (1991).
 14. R. K. Bose, *Copyright Issues in Multimedia*, SRI International Business Intelligence Program D92-1692, Menlo Park, Calif., 1992, p. 12.
 15. W. Peryman, *J. Library Admin.* **15**(1–2), 81 (1991).
 16. M. Fleming, ed., *Online Factbook*, Digital Information Group, Stamford, Conn., 1992.
 17. C. T. Meadow, *Database* **11**(5), 14 (1988).
 18. *Ibid.*, p. 23.
 19. R. K. Summit, *Online* **11**(1), 61 (1987).
 20. M. E. Williams, in K. Y. Marcaccio, ed., *Gale Directory of Databases*, Gale Research Institute, Detroit, 1993, pp. xvii–xxvii.
 21. S. E. Arnold, *Online* **13**(2), 6 (1989).
 22. P. P. Massa, *Bull. Am. Soc. Info. Sci.* **17**(6), 8 (1991).
 23. R. Brody, *Online* **17**(3), 66 (1993).
 24. T. B. Chadwick, *Online* **13**(1), 26 (1989).
 25. N. Garman, *Online* **13**(1), 6 (1989).
 26. R. Basch, *Database* **14**(5), 13 (1991).
 27. M. O'Leary, *Database* **13**(2), 15 (1990).
 28. J. Thompson, *Online* **13**(6), 11 (1989).
 29. H. Pemberton, *Online* **16**(2), 12 (1992).
 30. S. N. Björner, *Online* **14**(4), 90 (1990).
 31. M. O'Leary, *Online* **16**(3), 29 (1992).
 32. R. J. Massie, *The American Chemical Society and Dialog Information Services Settle Litigation*, press release, Chemical Abstract Service, Columbus, Ohio, Oct. 29, 1993.
 33. D. T. Hawkins, *Online* **17**(4), 98 (1993).
 34. M. O'Leary, *Online* **17**(1), 34 (1993).
 35. P. Lane, *Info. Today* **9**(11), 1 (1992).
 36. A. J. Van Camp, *Online* **16**(2), 102 (1992).
 37. A. Novallo, ed., *1994 Information Industry Directory*, 14th ed., Vol. 1, Gale Research Inc., Detroit, 1994.
 38. *The CIS (Chemical Information System): An Overview*, Chemical Information Systems, Towson, Md., 1993, p. 1.
 39. L. Eichler and J. Newland, *Database* **16**(3), 48 (1993).
 40. C. Wilson, ed., *The Chronolog* **21**(4), 91 (1993).
 41. *DLIALOG Database Catalogue*, Dialog Information Services, Inc., Palo Alto, Calif., 1993, p. 3.
 42. *ORBIT Database Catalogue*, Info Pro Technologies, 1993, McLean, Va., p. 1.
 43. *Questel Database Catalog-1993*, Questel, Inc., Arlington, Va., 1993, p. i.
 44. *July 1993 STN Database Catalog*, Chemical Abstracts Service, Columbus, Ohio, 1993, p. 4.
 45. B. D. Christie, B. A. Leland, and J. G. Nourse, *J. Chem. Inf. Comput. Sci.* **33**(4), 545 (1993).
 46. A. J. Gushurst, and co-workers, *J. Chem. Inf. Comput. Sci.* **31**(2), 447 (1991).
 47. M. G. Hicks and C. Jochum, *Anal. Chim. Acta* **235**, 87 (1990).
 48. R. Attias and J. -E. Buboïs, *J. Chem. Inf. Comput. Sci.* **30**(1), 2 (1990).
 49. J. M. Barnard, *J. Chem. Inf. Comput. Sci.* **33**(4), 532 (1993).
 50. M. G. Hicks and C. Jochum, *J. Chem. Inf. Comput. Sci.* **30**(2), 191 (1990).
 51. A. Barth, in S. R. Heller, ed., *ACS Symposium Series 436*, American Chemical Society, Washington, D.C., 1990, pp. 24–41.
 52. *Beilstein and Gmelin on STN*, Springer-Verlag, Heidelberg, Germany, 1993.
 53. P. G. Dittmar, *J. Chem. Inf. Comput. Sci.* **23**(3), 93 (1983).
 54. J. R. McDaniel and A. E. Fein, in W. A. Warr, ed., *ACS Symposium Series 341*, American Chemical Society, Washington, D.C., 1987, p. 62.
 55. Beilstein-Institut für Literatur der Organischen Chemie, Germany.
 56. Gmelin-Institut für Anorganische Chemie und Grenzgebiet, Germany.
 57. MPD Network, Columbus, Ohio.
 58. Technical data, Technical Databases Services, Inc. (TDS)-Numerica, New York.
 59. S. E. Jakes, and co-workers, *J. Molec. Graphics*, **5**(1), 41 (1987).
 60. F. H. Allen, M. J. Doyle, and R. Taylor, *Acta Crystallog., Section B*, **B47**(1), 50 (1991).
 61. R. A. Engh and R. Huber, *Acta Crystallog., Section A*, **A47**(4), 392 (1991).
 62. C. Craver, ed., *The Coblenz Society Desk Book of Infrared Spectra*, The Coblenz Society, Inc., Kirkwood, Mo., 1977.
 63. A. N. Davies, H. Hillig, and M. Linshcid, *Proceedings of the Scientific Computing and Automation (Europe) Conference*, 1990, p. 445.
 64. E. Pretsch, M. Farkas, and A. First, *Proc. Int. CODATA Conf.*, 11th ed., 1988, p. 176.
 65. W. A. Warr, *Chemometrics and Intell. Lab. Sys.* **10**, 279 (1991).
 66. S. G. Lias, *J. Res. Natl. Inst. Stand. Technol.* **94**(1), 25 (1989).
 67. E. Kono, A. Hidetsugu, and S. Sasaki, *Jobo Kagaku Toronkai, Kozo Kassei Soka Shinpojumu Koen Yoshishu* **13–18**, 149 (1990).
 68. M. G. Weller, in H. Collier, ed., *Recent Advances in Chemical Information*, Royal Society of Chemistry, London, 1992, pp. 197–208.
 69. S. R. Heller, *Chem. Int.* **13**(6), 235 (1991).
 70. O. Yamamoto, K. Hayamizu, and M. Yanagisawa, *Anal. Sci.* **5**(2), 141 (1989).
 71. O. Yamamoto, K. Hayamizu, and M. Yanagisawa, *Jobo Kagaku Toronkai, Kozo Kassei Soka Shinpojumu Koen Yoshishu* **13–18**, 53 (1990).
 72. K. S. Lebedev, E. A. Otmakhova, and I. V. Gritsenko, *Izv. Sib. Otd. Akad. Nauk SSSR, Ser. Khim. Nauk* **5**, 73 (1990).
 73. D. E. Meyer, W. A. Warr, and R. A. Love, eds., *Chemical Structure Software for Personal Computers*, American Chemical Society, Washington, D.C., 1988.

74. J. L. Markley, and co-workers, in M. Ikehara, ed., *Protein Eng. Proc. Int. Conf. Protein Eng., 2nd*, 1989, Japan Scientific Society Press, Tokyo, 1990, p. 285.
75. P. R. Griffiths and C. L. Wilkins, *Appl. Spectros.* **42**, 538 (1988).
76. D. D. Speck, R. Venkataraghavan, and F. W. McLafferty, *Org. Mass Spectrom.* **13**, 209 (1978).
77. F. W. McLafferty, and co-workers, *Int. J. Mass Spectrom. Ion Phys.* **47**, 317 (1983).
78. D. T. Terwilliger, and co-workers, *Biomed. Environ. Mass Spectrom.* **14**, 263 (1987).
79. *CIS (Chemical Information System): An Overview*, Chemical Information Systems, Inc., Towson, Md., 1993.
80. K. Hayamizu, *Kagaku to Sofutouwa* **14**(3), 191 (1992).
81. K. Tanabe, K. Hayamizu, and S. Ono, *Anal. Sci.* **7**, 711 (1991; Suppl., *Proc. Int. Congr. Anal. Sci.*, Pt. 1).
82. J. R. Rumble, Jr., D. M. Bickham, and C. J. Powell, *Surf. Interface Anal.* **19**(1–12), 241 (1992).
83. H. Chihara and K. Mano, *Z. Naturforsch., A: Phys. Sci.* **47**(1–2), 446 (1992).
84. K. Okano and A. Abe, *Jobo Kanri* **31**(1), 56 (1988).
85. G. H. Wood, J. R. Rogers, and S. R. Gough, *J. Chem. Inf. Comput. Sci.* **29**, 118 (1989).
86. P. E. Blower, in P. Willett, ed., *Modern Approaches to Chemical Reaction Searching*, Gower, Aldershot, U.K., 1986, p. 146.
87. *Chem. Pharm. Bull.* **41**(11), 1906–1909 (1993).
88. *REACCS Information Management Reference Manual*, revision 8.0, Molecular Design Ltd., San Leandro, Calif.
89. W. T. Wipke and T. M. Dyott, *J. Am. Chem. Soc.* **96**, 4825 (1974).
90. E. Aberdeen, *Group Market Viewpoint* **4**(11) (Nov. 12, 1991).
91. J. T. Butler, *LawLibraryJ.* **84**, 121 (1992).
92. E. S. Simmons, *Database* **8**(1), 49 (1985).
93. E. S. Simmons, *J. Chem. Inf. Comput. Sci.* **24**(1), 10 (1984).
94. N. Lambert, *Database* **10**(6), 46 (1987).
95. S. M. Kaback, **World Pat. Info.** **11**(2), 95 (1989).
96. E. S. Simmons, *Online* **10**(4), 51 (1986).
97. M. P. O'Hara and C. Pagis, *J. Chem. Inf. Comput. Sci.* **31**(1), 59 (1991).
98. J. Lucas, **World Pat. Info.** **14**(3), 167 (1992).
99. K. A. Cloutier, *J. Chem. Inf. Comput. Sci.* **31**(1), 40 (1991).
100. P. G. Alston and F. W. Stoss, *Database* **15**(4), 17 (1992).
101. T. Ebe, K. A. Sanderson, and P. S. Wilson, *J. Chem. Inf. Comput. Sci.* **31**(1), 31 (1991).
102. R. R. Freeman and M. F. Smith, in, M. E. Williams, ed., *Annual Review of Information Science and Technology (ARIST)*, Vol. 21, Elsevier Science Publishers, New York, 1986.
103. R. Basch, *Database* **16**(6), 47 (1993).
104. M. J. Cronin, *Database* **16**(6), 47 (1993).
105. C. B. Tilly, in, M. E. Williams, ed., *Annual Review of Information Science and Technology (ARIST)*, Vol. 25, Elsevier Science Publishers, New York, 1990.
106. H. M. Kissman and P. Wexler, in, M. E. Williams, ed., *Annual Review of Information Science and Technology (ARIST)*, Vol. 18, Elsevier Science Publishers, New York, 1983.
107. P. Hernon and C. R. McClure, in, M. E. Williams, ed., *Annual Review of Information Science and Technology (ARIST)*, Vol. 28, Elsevier Science Publishers, New York, 1993.
108. E. N. Mailloux, in M. E. Williams, ed., *Annual Review of Information Science and Technology (ARIST)*, Vol. 24, Elsevier Science Publishers, New York, 1989.
109. EPA, *Information Systems Inventory*, NTIS kit order no. PB-94-500501, Environmental Protection Agency, Washington, D.C., 1993.
110. EPA, Access EPA, EPA IMSD-91-100, GPO order no. 055-000-00378-5 or NTIS PB91-151563, Environmental Protection Agency, Washington, D.C., 1993.
111. EPA, *Databases Accessible Through EPA Libraries*, Environmental Protection Agency, Washington, D.C. (in press) (1994).
112. V. E. Hample, D. P. Grubb, and A. Moulik, in J. R. Williams, N. A. Gocken, and J. E. Morral, eds., *Computerized Metallurgical Databases*, Metallurgical Society, Warrendale, Pa., 1988, p. 19.
113. K. D. Schwarz, *Gov. Comp. News* **11**(2), 41 (1992).
114. L. F. Lunin, in M. E. Williams, ed., *Annual Review of Information Science and Technology*, Vol. 22, Elsevier Science Publishers, Amsterdam, 1987, p. 179.
115. E. K. Brumm, in M. E. Williams, ed., *Annual Review of Information Science and Technology*, Vol. 26, Learned Information, Inc., Medford, N.J., 1991, p. 197.
116. L. Krumenaker, *Science* **260**, 1066 (1993).
117. K. Y. Marcaccio, ed., *Gale Directory of Databases*, Vol. 2, Gale Research Inc., Detroit, Mich., 1993.
118. K. M. Ginther-Webster, *AI Rev. Prod., Serv., Res.* **3**, 25 (1990).
119. G. A. Story, *Computer* **25**, 17 (1992).
120. R. Fidel, *Special Libraries* **83**(1), 1 (1992).
121. B. R. Boyce and J. P. McLain, *J. Am. Soc. Info. Sci.* **40**(4), 273 (1989).
122. F. W. Lancaster, C. Elliker, and C. T. Harkness, *Ann. Rev. Info. Sci. Technol.* **24**, 35 (1989).
123. S. C. Biswas and F. Smith, *Library & Info. Sci. Res. II*, (2), 109 (1989).
124. P. Willett, *Document Retrieval Systems*, (1988).
125. A. Hopkinson, *9th International Online Information Meeting*, 295–304 (1985).
126. T. E. Wolff, *Database* **15**(2), 34 (1992).
127. C. Moon, *Int. J. Info. Manage.* **8**(4), 265 (1988).
128. D. Cheney and G. Jenks, *Library Software Rev.* **7**(6), 411 (1988).
129. G. Lundeen, *Database* **12**(3), 36 (1989).
130. S. Stigleman, *Database* **15**(5), 50 (1992).
131. A. Raeder, *Database* **14**(5), 67 (1991).
132. T. Hanson, *Library Micromation News*, (33), 16 (1991).
133. K. Watterson, *Data Based Advisor II*, (6), 45 (1993).
134. E. Perez, *Database* **15**(6), 45 (1992).
135. G. W. Lundeen and C. Tenopir, *Database* **15**(4), 51 (1992).

General References

- G. Anderson, *Acquisitions Librarian* **6**, 3 (1991).
- A. R. Haygarth-Jackson, *Serials '85. Proceedings of the U. K. Serials Group Conference*, 29–47 (1985).
- S. James, *Library Rev.* **39**(4), 21 (1990).
- H. Martin, *Law Library J.* **82**(1), 129 (1990).
- D. Tyckson, *Reference Librarian* **34**, 37 (1991).

Du Pont Co., Scientific Computing Division, Experimental Station, Wilmington, Del., 1988.

L. Katz, *Rolling Stone* **77** 25–36 (Apr. 15, 1993).

E. Krol, *The Whole Internet*, O'Reilly & Associates, Inc., Sebastopol, Calif., 1992.

C. A. Lynch and C. M. Preston, in M. E. Williams, *Annual Review of Information Science and Technology*, Vol. 24, Learned Information, Medford, N.J., 1989, pp. 263–312, 140 references.

Online files from Sura.Net.

M. Rapaport, *Online* **15**(3), 33 (1991).

H. N. Tillman, *Special List* **16**(3), 1,3 (1993).

B. Brain, *Australian Academic and Research Libraries* **22**(3), 167 (1991).

M. K. Duggan, *Online* **15**(3), 20 (1991).

J. Garrett, *Proceedings of the 12th National Online Meeting, 1991*, Learned Information, Medford, N.J., 1991, pp. 99–107.

L. Garson, *J. Chem. Inf. Comput. Sci.* **24**(3), 119 (1984).

M. Hattery, *Info. Retrieval & Library Automation* **28**(10), 1 (1993).

M. Jensen, *Serials Rev.* **18**(1–2), 62 (1992).

J. Ogburn, *Library Acquisitions, Practice and Theory* **14**(3), 257 (1990).

M. K. Booker, *Encycl. Materials Sci. and Engineering* **2**, 796–800 (1986).

N. Herrlich and J. Wierer, *Kunstst. Ger. Plast.* **79**(10), 106–109 (1989).

F. C. Allan and W. R. Ferrell, *Database* **12**, 50–58 (1989).

J. G. Kaufman, *Engineering with Computers* **4**, 75–85 (1988).

Third Chemical Congress of North America, American Chemical Society, June, 1988, Toronto, Canada, Herman Skolnik Award Symposium, Scientific Numerical Databases, Present and Future.

Association for Information and Image Management (AIIM), *Information & Image Management: The State of the Industry 1993*, Association for Information and Image Management, Silver Spring, Md., 1993.

P. Jasco, *CD-ROM Software, Dataware and Hardware: Evaluation, Selection and Installation*, Libraries Unlimited, Englewood Cliffs, Colo., 1992.

J. Langlois, *CD-ROM 1992L An Annotated Bibliography of Resources*, Meckler Corp., Westport, Conn., 1992.

G. Tapper and K. Tombs, *Admissibility of Document Imaging Systems*, Meckler Corp., Westport, Conn., 1992.

NERAC, bibliography, *Full Text Database Searching*, pub. no. PB92-860469, Tolland, Conn., 1992.

NERAC, bibliography, *Database Design*, pub. no. PB92-861665, Tolland, Conn., 1992.

Cynthia S. Barcelon-Yang
Evelyn L. Brownlee
Emmett D. Calhoun
Bruno A. Caputo
Charles C. Cumbo
Joseph P. Daniszewski
DouglasA. Eckel
Kenneth H. Glaspey
Darlyn C. Green-Kocher
Marianne B. Gruber
Margaret M. Isselmann
Thomas C. Johns
Alicia P. King
David M. Krentz
Florence H. Kvalnes
Luray M. Minkiewicz
Behrooz Nazer
Angela K. G. Parsons
Carol R. Perrotto
Rita D. Ratliff
Jeanette C. Sikes
Amie H. Webster
Du Pont Company

INFORMATION STORAGE MATERIALS

Optical,
Magnetic,

OPTICAL

A most important element in computer technology is data storage. Progress in microelectronics, therefore, is directly linked to progress in data storage, that is the ability to store large amounts of information in the smallest possible space, irreversibly or preferably reversibly.

For data storage two types of memory are distinguished: (1) main memory with moderate capacity (currently 4×10^6 bit in each element) and extremely short access time ($<10^{-7}$ s), and (2) mass memory with very high capacity ($>10^7$ bit in each element) and moderate access time ($\sim 10^{-2}$ s).

Main memories almost exclusively consist of semiconductors on a silicon basis in complementary metal oxide semiconductor technology (CMOS). The most important types are the pure read only memory (ROM) and the write read memory (RAM random access memory), which is available as S-RAM (static RAM) or as D-RAM (dynamic RAM).

The most important mass memories use magnetic media in the form of magnetic tapes or disks (floppy disk and hard disk). Laser addressed optical mass memories are of increasing commercial importance.

Polymers are only marginally important in main memories of semiconductor technology, except for polymeric resist films used for chip production. For optical mass memories, however, they are important or even indispensable, being used as substrate material (in WORM, EOD) or for both substrate material and the memory layer (in CD-ROM). Peripheral uses of polymers in the manufacturing process of optical storage media are, eg, as binder for dye-in-polymer layers or as surfacing layers, protective overcoatings, uv-resist films, photopolymerization lacquers for replication, etc.

Photopolymers and photothermoplasts are mentioned only in connection with holographic data storage (see HOLOGRAPHY). The classical method of optical data storage in silver halide films (photographic film, microfiche technique) is not discussed (see PHOTOGRAPHY).

Classification

In general, the commercially used optical data storage media deposit the information on disks or cards (two-dimensional data deposition, Table 1). Data storage systems, which store data in three and more dimensions are being developed.

Table 1. Methods of Two-Dimensional Data Storage on Disks

Symbol	Typical properties	Examples	Application of polymers
CD-ROM	not erasable not rewritable	technology identical with audio compact disk (CD-DA)	substrate and information layer from polycarbonate (PC)
WORM	writable not erasable	polymeric or glassy substrates with metal or alloy layer, dye-in polymer film	substrate $\theta < 51.4$ in. ^a : PC $\theta > 51.4$ in.: glass, partly PC
EOD	erasable rewritable	polymeric or glassy substrates with magneto-optical recording layer (MOR); phase change recording layer (PCR); photochromic dyestuff	substrate $\theta < 51.4$ in.: PC, partly glass $\theta > 51.4$ in.: not on the market

^a 51.4 in. is a standard disk size.

Depending on the method of data read-out, respectively read-in read-out, two systems are distinguished: mechano-optical systems with usually disk-shaped media (optical disks), and purely optical systems with card-shaped media without moving parts (optical memory cards).

The disk-shaped optical data storage media can be differentiated into three classes: (1) CD-ROM (compact disk-read only memory) in which data are embossed by the manufacturer. The lay-out of the disks and information storage are analogous to the well-known audio compact disk. (2) In WORM (write once, read many times), data are written by the user, but are not erasable or rewritable. (3) In EOD (erasable optical disk), data are written by the user and can be erased or overwritten. EOD was formerly known as E-DRAW (erasable-direct read after write). In Table 1 the different types of disk-shaped optical media for two-dimensional data storage are listed according to their construction, type, fundamental characteristics, examples of their implementation and use of polymers. Multidimensional data storage systems being currently developed are holographic data storage in which information is stored as a diffraction or as an absorption hologram or photochemical, respectively, photophysical hole burning (PHB). In holographic methods polymers are used as both matrix material and as storage media (see HOLOGRAPHY). In PHB polymers are used as matrix material.

Planar Data Deposition (2-D Storage)

DATA STORAGE WITH PRERECORDED INFORMATION

CD-ROM disks are nearly identical to the well-known compact disk-digital audio (CD-DA; short CD). The information on a CD-ROM is stamped in the form of clearly defined pits on the disk surface during the disk's manufacture, using injection molding or injection stamping techniques. A metal stamper transfers the digital information to the disk's surface.

The digital information is represented by the position and length of microscopic pits on the surface of a CD-ROM that are arranged in a spiral track. A CD-ROM of 120 mm (4.75 in.) diameter has a gross capacity (unformatted) of about 600 MByte and a net capacity (formatted) of 540 MByte

(ISO-Norm 9660), respectively. This is equivalent to 200,000 pages of text. The access time is between 200 and 600 ms. The data transfer rate of a standard audio CD player is 144 KB s, but dedicated CD-ROM drives can transfer data at up to 300 KB s by doubling the rotational speed of the disk.

Figure 1a schematically depicts a cut-out of a CD-ROM sector (1); the sketch shows the track-to-track distance (1.6 μm) and the width of a pit (about 0.6 μm). The pit length varies between 0.8 and 3.3 μm. The pit depth is 0.11–0.12 μm. Figure 1b shows the structure of the spiral grooves for tracking that are imprinted into the substrate material of CD-DA during manufacture. To read the stored digital information, low energy laser diodes (about 7 mW) are employed with emission wavelengths from 780 to 840 nm. As shown in Figure 2, the focused laser beam is reflected by the sputtered aluminum cover layer and received by a photodetector. In the process, the laser beam (ϕ (diameter) about 1 μ) is scattered in the pit area and weakened by interference, respectively, but reflected optimally in the flat areas in between. Thus, the digital data, which is coded with certain algorithms, can be read. CDs use bit-edge detection, ie, the high to low transition itself is responsible for the signal at the photodetector, registered as a 1 or as a 0, respectively.

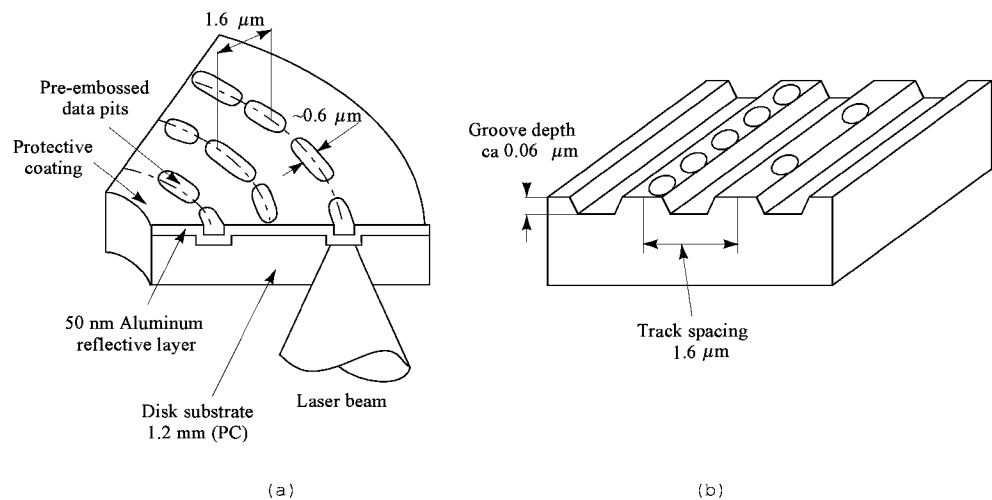


Fig. 1. Compact disk geometry: (a) CD-ROM, sector detail (1); (b) CD-DA, geometry of the tracks (grooves).

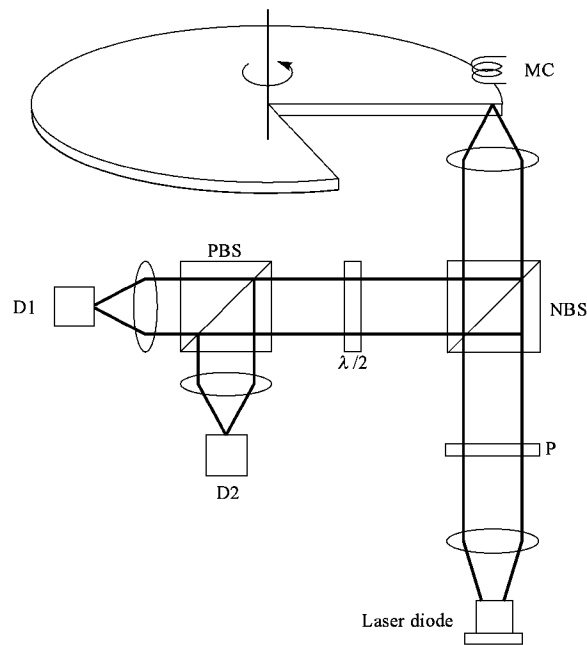


Fig. 2. A magnetooptical player (2). P = polarizer; MC = magnetic coil (for magnetooptical writing); NBS = neutral beam splitter; PBS = polarizing beam splitter. D1, D2 = detectors for differential detection. The optical path also comprises tracking and focusing optics which are not shown here.

The physical CD format has been recognized as an industry standard: all CD-ROMs are of the same size, and data are stored in the same way in the same physical format. Details are put down in the so-called Yellow Book (Philips Sony).

For the logical format of a CD-ROM, a standard has also been agreed on (High Sierra Proposal, normed as ISO 9660; extension rock ridge), which governs the storage of files and enables the common operating systems (Apple HFS, MS-DOS, UNIX, VMS) to access them.

Besides the established audio CD and CD-ROM, there are other variants of optical storage disks with imprinted information which differ in the way the data are processed.

The CD-I (compact disk-interactive) is a low cost alternative to the CD-ROM for the entertainment industry. CD-I is a subset of the CD-ROM standard data format. It allows the digital storage of data, audio, and video information in a form that permits rapid interaction with a computer. CD-I is compatible to CD-ROM and to CD-AD; $\theta = 120$ mm. The definition of the CD-I format is put down in the Green Book (Philips Sony).

CD-XA (CD-extended architecture) is the standard for CD-ROMs for data storage and compressed audio recording (up to four hours music in HiFi-stereo quality). The definition of the CD-XA format is also put down in the extended Yellow Book.

CD-R (CD-recordable) is a writable, nonerasable disk, also called CD-WORM or CD-WO (CD-write once). Permanent marks are produced by a focused laser beam. The definition of the CD-R format and of the erasable–rewritable EOD MO-R format is put down in the Orange Book (Philips Sony).

CD-V (compact disk-video) is premolded for both video and digital music; $\theta = 305$ mm or $\theta = 203$ mm. There is also a videoclip version ($\theta = 120$ mm) for 22 min CD audio and 5 min of video information. CD-VEP (video extended play) $\theta = 200$ mm, is recorded on both sides, whereas for CD-VLP (video long play) $\theta = 200$ mm, one- or two-sided recording can be selected; max. 60 min video side.

Not in the class of digital recording storage media are video CDs with diameters of 200 and 300 mm, respectively. These are storage disks for recording the analogue signal of videos and are of air-sandwich construction with two poly(methyl methacrylate) disks.

WRITABLE, NONERASABLE DATA STORAGE

In many applications, data storage systems are required which enable the user to write data and text as well as in some cases digitized graphics and pictures. They should allow fast access to the stored information at all times. The information itself, however, may not be changed, erased, or overwritten. Examples for these applications are mainly office files, especially those which require mandatory storage, protection against manipulation, and forgery-proof documents, eg, expense statements, payment and salary files, bank statements, employee files, or production instructions. In general, these storage systems could substitute for filing of documents on paper or archiving on microfiche.

The conditions of these target applications are fulfilled by WORM-disks. Apart from data compression on a small volume, WORM filing systems offer the advantage of fast access from the workplace at all times, including a simplified document search and retrieval strategy.

A WORM disk generally consists of two polycarbonate disks with 130 mm diameter in sandwich construction. The storage capacity of a WORM disk varies depending on the disk diameter, the manufacturer, the mechanism, and disk formatting. Currently, WORM disks of 5.25 in. format are offered with 2×235 MByte or 2×470 MByte capacity (double-sided), and of 12 and 14 in. format with capacities ranging from 1 to 4.1 GByte (double-sided). A breakthrough in WORM applications is the introduction of a WORM disk in CD format (so-called CD-WORM or CD-R; R = recordable), offering the possibility of using CD-WORM disks in customary, new generation CD players (multimedia operation).

A special implementation of the CD-R disk is the Photo-CD by Kodak which is a 5.25 in. WORM disk employing the dye-in-polymer principle for storage of up to 100 slides pictures on a CD (after data compression) with the possibility of interactive picture processing.

WORM disks must fulfill the following requirements (3): high storage capacity ($\sim 2 \times 10^7$ bit cm^2), short access time ($< 10^{-1}$ s), read back with high signal-to-noise ratio > 47 dB, long shelf life of the information (> 20 years), low storage costs ($< 10^{-7}$ \$/bit), low error rate (BER = bit error rate $< 10^{-12}$ bit, after error correction), and high reliability (MTBF = mean time between failure > 2000 h).

Writing Techniques. WORM disks differ depending on their data writing techniques, which can be divided into three classes (3): ablative writing, bubble forming, and phase change.

Ablative Writing. This technique involves burning or melting of pits. Holes may be burned into thin dye layers, preferably organic pigments. Burning of holes into thin metal alloy layers, eg, Te or Bi films is another method. Holes are burned into polymeric active dye layers, which contain light-absorbing additives tuned to the laser wavelength used (organic dyes or pigments, inorganic pigments, carbon black, etc). Especially good effects are achieved with a strong interaction between dyes and the polymeric binder, the dye-in-polymer (DIP) concept (4) (Fig. 3a). The melting of pits in connection with heat-induced diffusion of an organic dye into the polymeric binder is the laser-induced dye amplification (LIDA) technique (5) (Fig. 3b).

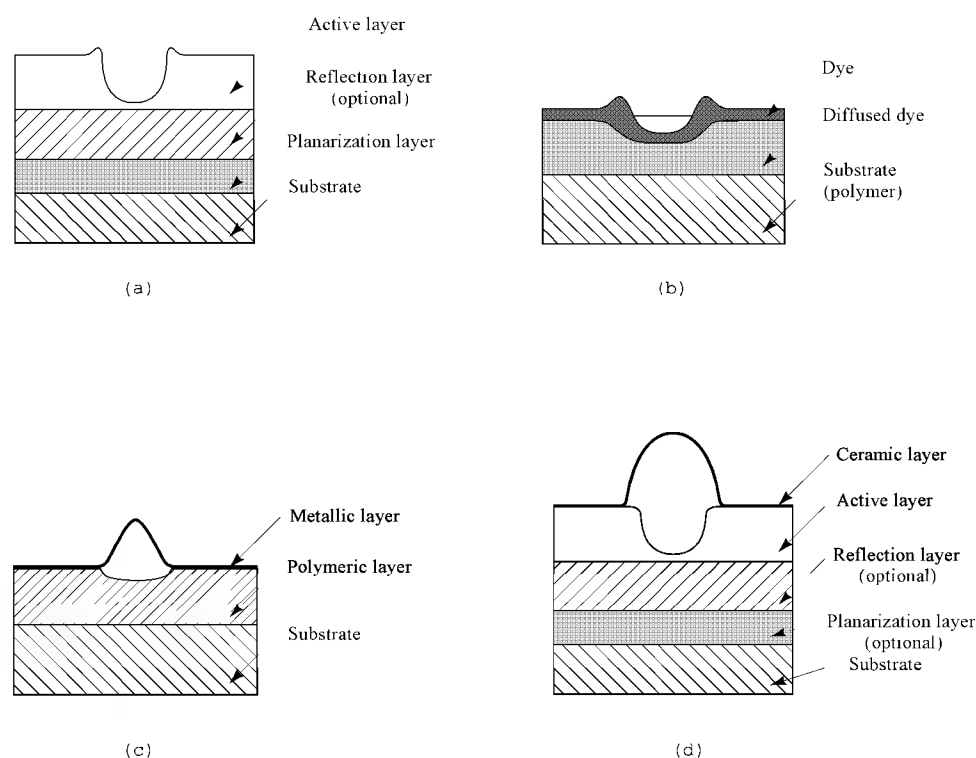


Fig. 3. WORM media: data writing techniques and typical layer construction (3): (a) ablative writing, DIP concept (4); (b) ablative writing, LIDA technique (5); (c) writing by bubble forming; (d) writing by bubble forming, layer arrangement after (6).

Writing by Bubble Forming. Bubble formation occurs under thin metal layers on polymeric substrate films, caused by local evaporation when hit by a focused laser beam (see Fig. 3c). Bubble formation occurs as in the DIP concept in dye-in-polymer films which are covered by a thin metal (mostly gold) or ceramic layer (6) (see Fig. 3d).

In variations of these first two methods, the information is written by creating holes/pits/bubbles of about $1 \mu\text{m}$ size with a laser (power 20–40 mW, impulse length about 50 ns). Reading is also done optically by laser (power about 0.5 mW) using the changed light reflection or scattering characteristics of these marked patterns. Here it is advantageous to improve the efficiency of optical absorption and to improve the signal contrast significantly by an additional thin, metallic reflecting layer beneath the memory layer. These are called tuned optical structures.

The mechanism of hole- or bubble-forming in metal or dye/polymer layers continues to be a subject of intensive investigation (7).

Writing by Phase Change. In an amorphous layer, crystalline marks are generated by local annealing with a focused laser beam.

Dyes for WORM-Disks. Regarding their memory layer, dye-in-polymer systems show advantages over metal layers in their higher stability, lower toxicity, lower heat conductivity, lower melting and sublimation temperature, and simpler manufacturing technique (substrate coating by sublimation or spincoating).

The following requirements need to be fulfilled by dyes or dye-in-polymer systems as active components in WORM-disks: high absorption capability at the wavelength of the write laser (wavelength 780–840 nm is low writing energy); defined threshold to avoid destructive reading; low heat conductivity parallel to the disk surface to yield focused pits, ie, high storage density, low toxicity, good solubility in solvents (eg, pentanol, hexanol) which do not attack the disk material (generally polycarbonate); and good film forming, ie, low cost manufacturing technique.

The GaAs laser used as light source emits at about 820 nm. Thus dyes in the actual sense are not needed; rather, ir-absorbers for the spectrum between 750–850 nm; little experience is available on this class of dyes, especially as far as their stability is concerned, although much work has been done in this area. Also, infrared sensitive dyes and pigments, used in electrophotography, may be very suitable for WORM disks (8).

To yield high storage densities, layer thicknesses in the order of the focused laser beam are necessary in the storage medium, measuring about $1 \mu\text{m}$. In commercial WORM disks based on pure dye layers, layer thicknesses are even lower at about $0.1 \mu\text{m}$. With thicknesses as low as these, it becomes

difficult to ensure sufficient light absorption, so the dyes need to possess a high molecular extinction. A low heat conductivity of the storage layer is of equal importance; only layers with low energy dissipation yield small focused pits already at low write energies, and with that a high storage density.

The dyes used can be classified in four groups (3) as follows.

Methine Dyes. In this dye class, special importance has been gained by zwitterionic hydroxy squarylium dyes (9) (Fig. 4a) (10), the cationic dye SQS (squarylium core with thiopyrylium end group (11) (Fig. 4b), and the cationic polymethine tetra(dimethylaminophenyl)pentamethineperchlorate (TPMP) (11) (Fig. 4c) due to their favorable specific combination of characteristics: high optical absorption at the laser wavelength, high signal-to-noise ratio on reading, low error rate, and reasonable production cost. As an example, the reflectance R and the transmittance T of a 75-nm thick SQS film is plotted as a function of wavelength λ in Figure 5 (11).

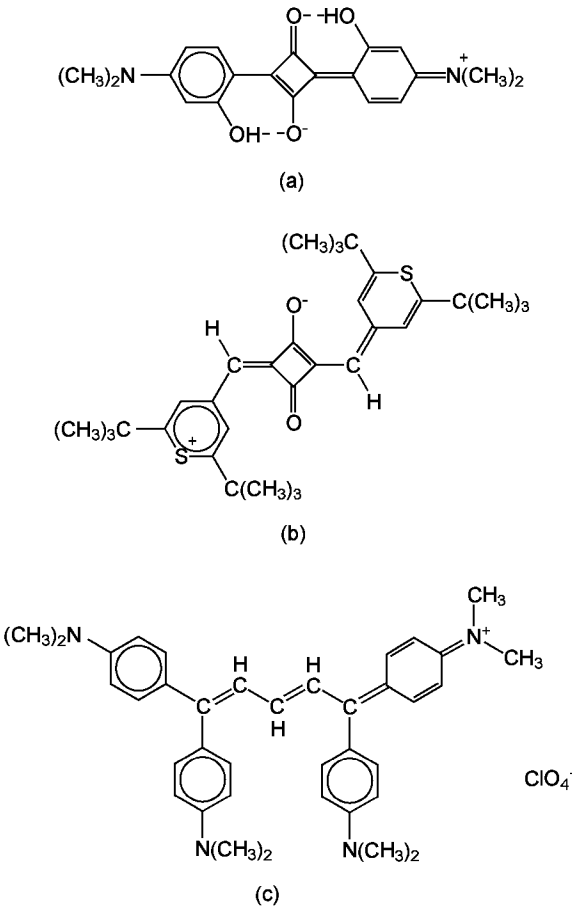


Fig. 4. Dyes for WORM media: (a) hydroxy squarylium (9,10); (b) SQS – squarylium core with thiopyrylium end group (11); (c) TPMP – tetra(dimethylaminophenyl)pentamethineperchlorate (11).

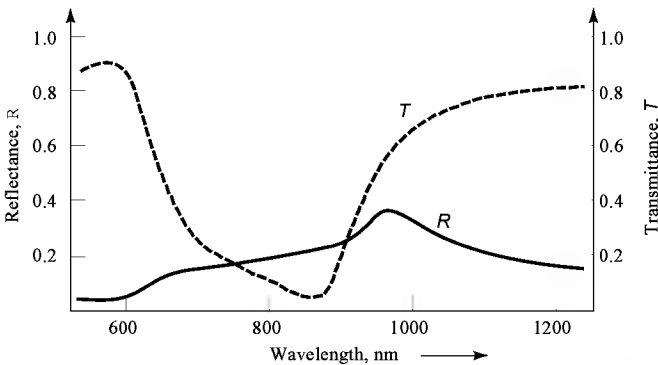


Fig. 5. Reflectance and transmittance spectra for a 75-nm thick SQS film (11).

In WORM disks, pentamethine and heptamethine are the materials of choice (see CYANINE DYES; POLYMETHINE DYES).

Naphthalocyanine Derivatives. These dyes, which contain a naphthalocyanine skeleton as chromophore, show high absorption in the area around 830 nm, possess low melting points, and small heat conductivity. A disadvantage of the unsubstituted naphthalocyanine is its low solubility in organic solvents which do not attack the polymeric substrate material (generally polycarbonate); substituents improve solubility (see PHTHALOCYANINE COMPOUNDS).

The basic structure of a naphthalocyanine dye can be seen in Figure 6 (12). By varying the central atom Y , the organic (polymeric) group Z_p at the central atom, and the substituents X_m and X_n , an adjustment to the desired property profile, especially improved solubility, can be achieved.

Naphthalocyanines with silicon as central atom evoke special interest, eg, the biaxially substituted silicon–naphthalocyanine (13) wherein the central Si bears two Z_p substituents and the rings are unsubstituted, ie, $X_m - X_n - H$.

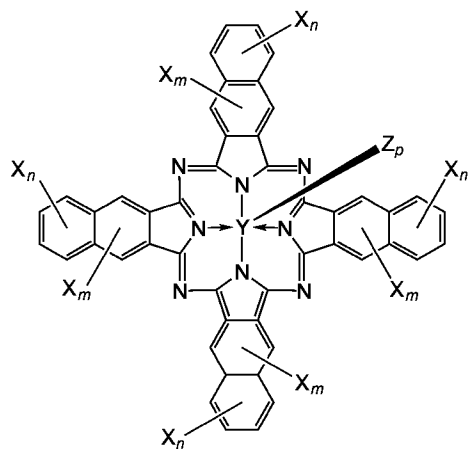
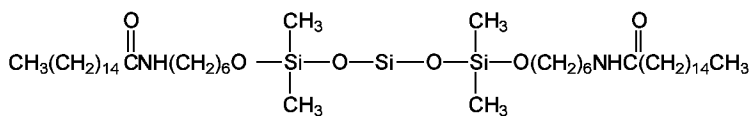


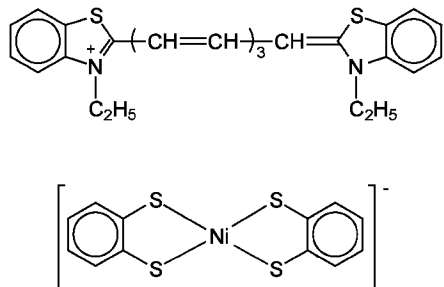
Fig. 6. Dyes for WORM media: phthalocyanine derivatives. The basic structure (12) of naphthalocyanine derivatives. Y = Si, Ge, Sn, Al, Ga, In, or a transition metal; Z_p = OR₁, OSiR₂R₃R₄, polymer. X_m and X_n represent substituents on the rings of the naphthalene system.



This dye shows high absorption at wavelengths between 780 and 830 nm, and a reflectivity of about 30%. This signal-to-noise ratio is about 50 dB with high consistency, even after 10¹⁰ read cycles (14). A disadvantage is the high expenditure for the synthesis of these derivatives due to low yield (15).

Naphthoquinone Derivatives. These dyes are generally not very colorfast and sometimes also fade. Anthraquinone derivatives have better properties (see DYES, ANTHRAQUINONE). New developments are based on compounds that combine the chemical structure of phenothiazine or phenoselenazine derivatives with benzoquinone or naphthoquinone units (16). These dyes show good chemical stability. Naphthoquinone derivatives have yet to find a practical application as dyes in WORM disks because of the high expenditure for synthesis due to low yields.

Metal Complexes. The importance of Ni complexes is based on their effectiveness as quenchers for singlet oxygen. Of disadvantage is their low colorfastness and their lower ir-reflectance compared to cyanine dyes (qv); therefore they are used in combination with suitable dyes. Numerous complexes are described in the literature, primarily tetrathiolate complexes of Pt or Ni, eg, dithiolatonickel complexes (3). Well known is the practical use of a combination of benzothiazole dyes with nickelthiol complexes in WORM disks (Ricoh, TDK) (17).



ERASABLE, REWRITABLE OPTICAL DATA STORAGE

The principal use of CD-ROM and WORM disks is essentially substitution of data storage on paper or microfiche. Conservative estimates number the worldwide use for data storage by paper at 91%, microfiche at 4%, and in electronic media at 5%, of which 4% are magnetic and 1% optical media (18). CD-ROM is being used as an electronic counterpart to print media; the WORM disk presents itself more and more as a substitute for paper to store archivable, forgery-proof documents.

Erasable optical disk (EOD) systems, on the other hand, are challenging classic magnetic media in some areas of application, primarily magnetic tape and the hard disk, but mostly optical media complement magnetic media.

In spite of the enormous advantages of erasable optical disks (higher storage density, trouble-free removable media, easy handling, low susceptibility to interference, low cost per bit), their market share of the whole mass storage market is relatively small as of the mid-1990s, contrary to earlier optimistic prognoses. This is less due to the specific disadvantages of the EOD compared to magnetic hard disks (longer access time, lower write-read speed), but primarily to a higher price.

High demands are made on erasable, rewritable optical data storage: high storage density (~2 × 10⁷ bit/cm²), short access time (<30 ms), reading with high signal-to-noise ratio (>47 dB), high data transfer rate (>0.7 MByte/s), high number of read-write cycles (>10⁷), long guaranteed shelf life of data (>10 years), low susceptibility to dirt and disturbances, trouble-free removable media, low cost per bit (<10⁻⁷ \$/bit), low bit error rate (BER < 10⁻¹², after correction), and high reliability (MTBF > 10⁶ h). EOD of the MO-R (magneto-optical recording) principle that have been on the market for several years, and the more recently introduced EOD of the PC-R (phase change-recording) principle, fulfill these requirements to a large degree, except for the long access time (20 to 80 ms) that should still be decreased. The data transfer rate can be raised, for example by increasing the rotation speed of the disk and/or the use of laser diode arrays.

Magneto-optical Recording. In a simplified way, a magneto-optical recording (MO-R) system can be regarded as a CD recorder using polarizing optics, a laser of controllable intensity, and a magnetic coil (see Fig. 2) (2). Currently, near ir-laser wavelengths of 780–820 nm are used, but red and blue recording which enables higher storage densities has been discussed. A disadvantage of MO storage technology in comparison to magnetic hard disks is the longer average access time: typically 30 ms, as compared to 10 ms. This is caused by the mechanical inertia of MO write-read heads. Because of the small track pitch, great demands are placed also on the actuator which controls head motion; to reduce the access time, intensive research is under way to produce lighter integrated optical heads.

As storage medium a (pregrooved) disk containing a hard magnetic layer is used where the information is stored in magnetic domain patterns. Currently used track pitches range from 1.35 to 1.6 μm. The principal advantages of MO recording in comparison to magnetic recording on hard disks are the removability of the disks and the immunity to head-crashes, the insensitivity against dirt, dust, fingerprints, etc (well known from audio CDs), which stem from the larger distance both of the optical head and the magnetic coil to the disk. The removability of the disks combined with high area densities leads to high total data capacities of MO recording systems. In 1994, 1.3 Gbyte storage capacity on 5.25 in. disks can be achieved for a track pitch of 1.35

μm . In the thermomagnetic write process (Fig. 7), a focused laser beam with a typical power of 10 mW locally heats the magnetic layer to its Curie temperature, where the spontaneous magnetization and the coercivity vanish. During cooling, the magnetization in the heated spot can be reversed by a small magnetic field (demagnetizing field of the layer itself and externally applied field), creating a magnetic domain. A cylindrical domain written in this way can be regarded as a frozen-in magnetic bubble and has correspondingly been treated both theoretically (19,20) and experimentally (21,22). Because of the high coercivity at room temperature, the data stability is excellent.

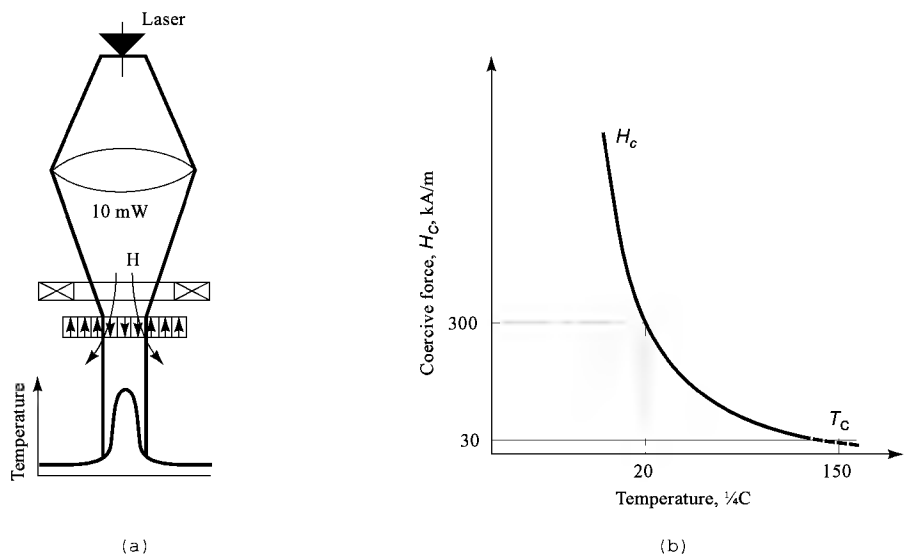


Fig. 7. Thermomagnetic recording. (a) A focused laser beam generates a thermal profile in the magnetic layer. (b) The coercive force in the layer is reduced and its magnetization can be reversed by a small magnetic field, here 30 kA m. At room temperature, the coercive force is high and the written domains are stable.

The read process utilizes the polar magneto-optical Kerr effect: the polarization plane of the reflected light is rotated clock- or counterclockwise by a perpendicular magnetization (up or down) in the film according to the domain pattern (Fig. 8). The minimum usable domain size is determined by the optical resolution during read-out. A typical size is 1 μm , which corresponds approximately to the diameter of the laser spot. For the materials used, the magnitude of the Kerr rotation is only 0.2 to 0.3°. Consequently, the read-out signals are fairly small. They can, nevertheless, be detected with a good signal-to-noise ratio (SNR), if the noise generated by the reflection and polarization reflections at the disk can be kept low. This is a necessary but not sufficient condition. The key for obtaining a high signal-to-noise ratio is the domain regularity. MO recording systems require a SNR >45 dB; optimized high performance disks achieve 60 dB.

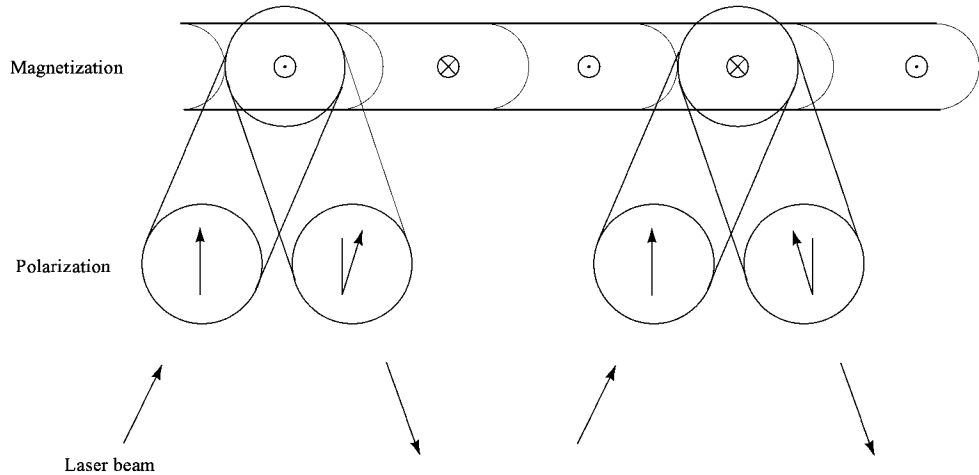


Fig. 8. Principle of the magneto-optical read-out of domain patterns by the polar Kerr effect. The polarization plane of the incoming laser beam is rotated clock- or counterclockwise according to the orientation (up or down) of the magnetic moments.

Two write methods are in use: magnetic field modulation (MFM) and laser modulation (LM) (2). With MFM, the whole track is heated by a laser beam of constant power and the magnetic field is switched according to the signal. Therefore, MFM is like a laser-assisted magnetic recording. This results in crescent-shaped domains like those shown in Figure 8. After writing, all old domains have been erased and only new information is on the track (direct overwrite, DOW). The data rate is limited by the inductance and heating of the magnetic coils. Practically achieved data rates are about 0.7 MByte s (23). Therefore, in order to make the coil smaller, there have been attempts to reduce the necessary write field; theoretical limits are discussed in Reference 24. One possibility is to decrease the distance to the MO layer by placing the coil on the layer side of the single-sided disk. Another possibility is to make the MO layer more field-sensitive, eg, by covering it with a soft magnetic capping layer (25,26).

With LM, the magnetic field is kept constant while the laser is switched according to the signal. Because only the laser power is modulated, possible data rates are higher than for MFM. They are limited by the present coding electronics to about 2 Mbyte s, the same as for magnetic hard disks. Therefore, LM is the preferred mode for data storage applications. The main drawback of this method is that only new domains are written but no old information is erased. Whole sectors have to be erased in a separate run before new information can be stored (no DOW). A solution to this problem is the use of multilayers.

From the write and read process sketched so far, some requirements for MO media can be derived: (1) a high perpendicular, uniaxial magnetic anisotropy K_u in order to enable readout with the polar Kerr effect; (2) a magneto-optically active layer with a sufficient figure of merit $R \cdot \theta_K$ where R is the reflectivity and θ_K the Kerr angle; (3) a Curie temperature between 400 and 600 K, the lower limit to enable stable domains at room temperature and the upper limit because of the limited laser power for writing.

The coercivity of a material is determined by the coefficient $K_u \approx 2M_i$ (27,28). If it is sufficiently high, demagnetizing effects during writing are avoided

and square hysteresis loops ensured. Therefore, it is advantageous that the magnetization of the MO layer is not too high between its Curie temperature and room temperature.

Materials for MO Media. The materials classes best suited for MO media are RE-TM alloys, Co–Pt multilayers, and ferrites. The quality of disks is mainly characterized by the signal-to-noise ratio (SNR).

Rare-Earth Transition-Metal Alloys. Amorphous thin films consisting of rare-earth–transition-metal (RE-TM) alloys were prepared in 1972 by coevaporation of Gd and Fe in the compositional range 15–94 atomic % Fe (29). In 1973 it was discovered that sputtered Gd–Co and Gd–Fe films contain striped and cylindrical domains indicating perpendicular anisotropy, and it was demonstrated that these films are suited as media for MO recording (30). Since then, RE-TM alloys have emerged as the material of choice for erasable optical recording. This preeminence results from a fundamental understanding of the basic magnetic material properties of RE-TM alloys as well as their deposition and optical recording properties and processes. An extensive review of the magnetism of amorphous RE-TM alloys can be found (31). The specific aspects relevant for recording are treated in References 32–39. A review (40) emphasizes among other things the domain dynamics and magnetization reversal in RE-TM films and RE-TM periodic multilayers.

RE-TM films can be prepared in a wide compositional range by electron beam evaporation, r-f (radio frequency) and d-c (direct current) diode, or magnetron sputtering (3). Magnetron sputtering is the method of choice for deposition, because it allows deposition at low argon pressure which gives optimum micromagnetic properties and high deposition rates for efficient production. The main deficiency of these materials is their corrosion due to the presence of rare-earth elements. Protective layers are therefore necessary.

The magnetic moments of the heavy RE elements (Gd, Tb, Dy, etc) are coupled antiparallel to the magnetic moments of the TM elements (Fe, Co, etc). The $RE_{1-x}TM_x$ alloys are therefore ferrimagnetic below their Curie temperature (T_C). The heavy TM moments form one magnetic sublattice and the RE moments the other one. In contrast, the light RE moments (eg, Nd, Pr) couple parallel to the moments of TM. The RE spin is always antiparallel to the TM spin, but for the light RE elements, the orbital momentum is coupled antiparallel to the spin and larger than the spin.

The temperature dependence of some key magnetic properties of a typical ferrimagnetic RE-TM composition is shown in Figure 9. The temperature at which the opposite magnetizations of the sublattices are equal in magnitude and the net magnetization is zero is called the compensation temperature (T_{comp}). At this temperature, the coercive field H_c needed to reverse the magnetization becomes infinite. In comparison, the MO effects at 800 nm wavelength (Kerr rotation angle θ_K), which are mainly due to the TM sublattice, are a continuous function of temperature. The presence of a compensation temperature can be used to control H_c and M_r . In the temperature range where the thermomagnetic switching process takes place, the maximum value of M_r above T_{comp} is approximately proportional to $T_C - T_{comp}$. A practical compensation temperature is found in a very narrow compositional range around 75 atomic % TM.

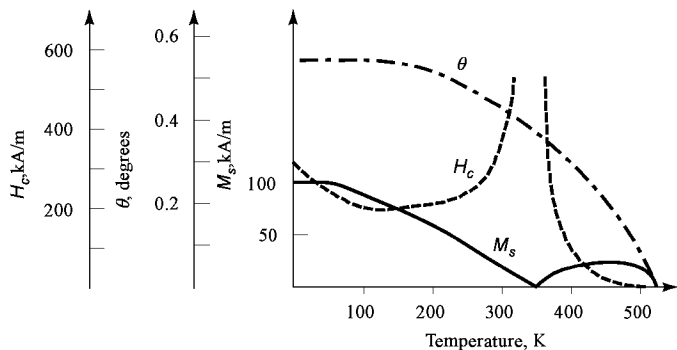


Fig. 9. Temperature dependences of the saturation magnetization M_s , the coercivity H_c and the Kerr rotation θ_K for a 50-nm $Gd_{0.24}Tb_{0.01}Fe_{0.75}$ film.

Figure 10 presents the Curie temperature (T_C) vs the TM-content (x) for Co- and Fe-based binary alloys. Alloying rare-earth elements with small amounts of transition metals ($x < 0.2$) leads to a decrease in Curie temperature. This is particularly obvious in the Gd–Co system where it corresponds to a nonmagnetic dilution similar to that of Cu (41,42). This indicates that TM atoms experience no exchange coupling unless they are surrounded by a minimum number j of other TM atoms. The critical number is $j = 5$ for Fe and $j = 7$ for Co. The steep increase of T_C for Co-based alloys with x about 0.7 is based on this effect.

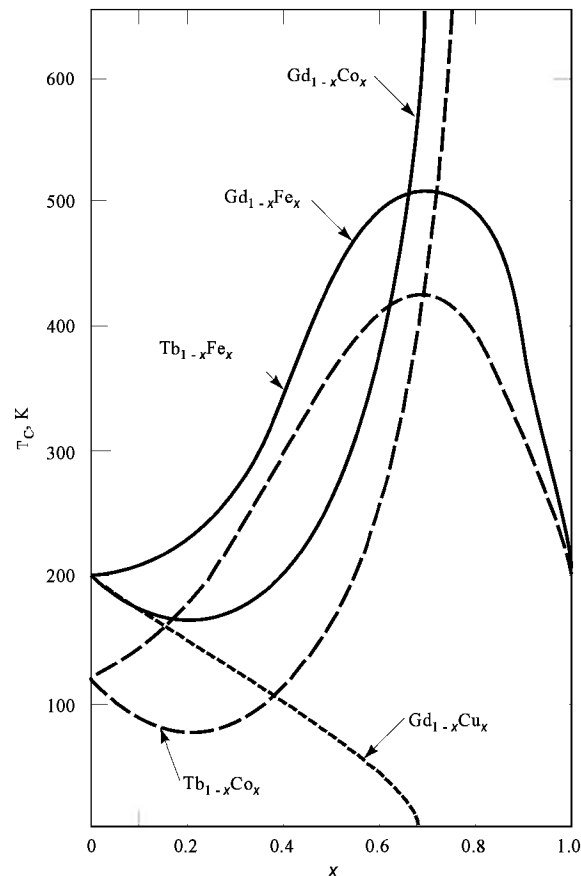


Fig. 10. Compositional variation of the Curie temperature (T_C) vs the TM content (x) for Co- and Fe-based binary RE-TM alloys. In the range of practical interest (about 75 atomic % TM), T_C is nearly independent of the Fe content, but strongly dependent on the Co content (about 7 K/atomic %).

For Fe-based alloys, a maximum in $T_C(x)$ can be found at $x = 0.7$. When x is increased further, the low T_C of amorphous Fe (approximately 200 K) is approached, due to the very strong influence of the structural disorder on the Fe-Fe exchange interaction leading to sperimagnetic order (43). The Curie temperature of Fe-based alloys can be increased by adding a small amount of Co. This effect allows a very fine tuning of the Curie temperature in the recording relevant region.

The magnetooptic spectra of amorphous RE-TM alloys are determined by contributions from both the TM and RE (44). In the ir wavelength region (used in today's MO-R), the MO effect stems mainly from the TM. In the visible and especially in the uv, a contribution of RE has to be taken into account. The RE contribution adds to the TM effect for the light RE, Nd and Pr, but reduces it for the heavy RE, Tb and Dy.

There have been many experimental investigations on the uniaxial (perpendicular) anisotropy constant K_u in amorphous RE-TM films. A synthesis of much of the existing data has been attempted in Reference 45. A pseudocrystalline or cluster model was proposed based on the idea that the short-range order in the amorphous state is similar to a relaxed crystalline one which in the compositional range around RE_{0.25}-TM_{0.75} consists of a layering of TM-planes and RE-rich planes. It is consistent with the observed coordination numbers in amorphous films. Though there is no direct evidence of such structural units, it can be said that the model is able to explain the known experimental results on the effects of composition, substrate temperature, annealing, and bombardment during sputter deposition on the resulting magnetic anisotropy.

Extended x-ray absorption fine structure measurements (EXAFS) have been performed to investigate the short-range structure of TbFe films (46). It is observed that there is an excess number of Fe-Fe and Tb-Tb pairs in the plane of the amorphous film and an excess number of Tb-Fe pairs perpendicular to film. The increase of K_u with the substrate temperature for samples prepared by evaporation is explained by a rearrangement of local absorbed atom configurations during the growth of the film (surface-induced texturing) (47).

In Table 2 the magnetic fundamentals of the three binary systems Gd-Fe, Tb-Fe, and Dy-Fe relevant for recording are listed, together with the parameter values required for recording (37,39). These alloys are ferrimagnetic, exhibiting for a certain compositional range a compensation temperature T_{comp} at which the net magnetization is zero. In order to get low magnetizations above room temperature, which is necessary to get a high T_C and to suppress subdomain formation due to demagnetizing effects, a compensation temperature around room temperature is favorable. This is obtained for about 75 atomic % Fe.

Table 2. Magnetic Fundamentals of Rare-Earth Transition-Metal Thin Films

Property	Gd _{1-x} Fe _x $x = 0.17$	Tb _{1-x} Fe _x $x = 0.75$	Dy _{1-x} Fe _x $x = 0.82$	Desired
compensation point at 295 K				
dT_{comp}/dx , ^a K/at. %	114	30	23	medium
anisotropy ^b K_u , 10 ⁵ J/m ³	-0.2	+6	+1	> +0.3
T_C , K	500	405	340	450
dT_C/dy , ^c K/at. % Co	7	7	10	

^a Change of compensation temperature with composition is dT_{comp}/dx .

^b Positive value of K_u indicates perpendicular anisotropy.

^c Increase of Curie temperature when adding Co is dT_C/dy .

None of the binary compounds with this composition is well matched to the needs of MO recording. Gd-Fe has too high a Curie temperature and has an in-plane anisotropy. T_C is too low for binary alloys such as Tb-Fe and Dy-Fe. Co-based alloys which exhibit a T_{comp} close to room temperature have

too high a Curie temperature. Practical MO RE-TM alloys are therefore ternary: GdTb-Fe, where the Tb introduces perpendicular anisotropy (48) increasing the coercivity, and decreasing T_C ; Tb-FeCo (49) and Dy-FeCo, where the Co increases T_C and the Kerr rotation angle, θ_K .

Each of the systems has an optimum distance T_C-T_{comp} for which the SNR has a maximum. For Tb-FeCo this is at about 220 K (50), corresponding to 23–27 atomic % Tb and for Dy-FeCo at about 120 K (51), corresponding to 25–30 atomic % Dy. The corrosion resistance of the compositions can be improved by adding a few atomic % of, eg, Cr, Ti, Nb, In, or Pt. These metallic additives have some impact on the magnetic properties of RE-TM alloys (45). K_u is reduced to some extent and, especially for 3d-additives, the TM moment is decreased. Therefore, the RE-TM composition must be reoptimized when dopants are added.

Pt/Co Periodic Multilayers. Since the discovery of perpendicular magnetic anisotropy in periodic Co-Pd multilayers (ML) (52), many investigations have been carried out on similar structures with alternating ultrathin Co-X layers with X = Pd (53,54), Pt (53–55), and Au (56). The elements Pd and Pt are magnetically polarizable by Co and the multilayers have very similar magnetic properties. For properly chosen thicknesses of the individual layers, the MLs exhibit perpendicular anisotropy and are therefore principally suited for MO recording. The big advantage of such systems is their high chemical stability, allowing simple stack designs without protective layers. In order to optimize the reflectivity, however, at least one dielectric layer between the substrate and the MO layer is needed. Co-Pt MLs are practically most important, because they exhibit the largest Kerr effect.

The magnetic properties of Co-Pt MLs depend on the thicknesses of the individual layers, t_{Pt} and t_{Co} (57). For $t_{Co} < 1.2$ nm, the axis of easy magnetization is perpendicular to the surface, as desired for MO recording. The maximum anisotropy is found for $t_{Co} = 0.4$ nm, corresponding to two atomic layers. Square hysteresis loops with 100% remanence are observed for $t_{Co} < 0.5$ nm and a total ML thickness of about 20 nm. A practically used configuration is, eg, 14 × (0.4 nm Co + 1.2 nm Pt). The information in these media is magnetically stored in about 30 layers of magnetic Co atoms.

The Kerr rotation angle (θ_K) of Co-Pt MLs increases in the ir and visible regions with increasing Co content. In the blue wavelength region it is larger than that of GdTb-Fe (Fig. 11) (58). The maximum of θ_K at 4.2 eV can be ascribed to a magnetooptic effect of Pt, where Pt is polarized by adjacent Co atoms (59). The Curie temperature decreases with increasing t_{Pt} . In Figure 11, sample D ($t_{Pt} = 1$ nm) has $T_C = 390^\circ\text{C}$ and sample A ($t_{Pt} = 1.94$ nm) has $T_C = 300^\circ\text{C}$. For MO media, a compromise between high Kerr rotation and low write power must therefore be found.

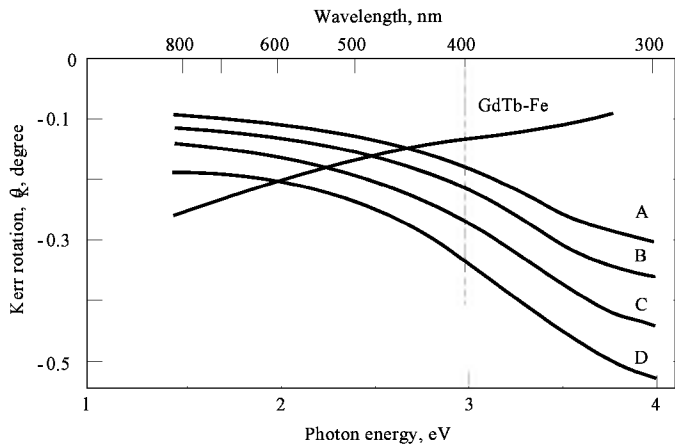


Fig. 11. The spectral dependence of the polar Kerr rotation for GdTb-Fe and for a series of Co-Pt multilayers with $t_{Co} = 0.36$ nm and with A, $t_{Pt} = 1.94$ nm; B, 1.63 nm; C, 1.28 nm; and D, 1 nm. Total thickness of all MLs is about 42 nm. Therefore, the Co content increases with decreasing t_{Pt} (adapted from Ref. 58).

Evaporated MLs exhibit a higher coercivity than Ar-sputtered films (sputtered at low pressures) and result also in the best recording results (60). With Ar-sputtering, the growing layer is exposed to a bombardment of reflected argon atoms with an energy of about 100 eV. This seems to enhance the surface diffusion, leading to sharp interfaces, but a distributed texture with twins. The energies for evaporation (about 0.2 eV) and Xe sputtering (about 1 eV) are lower, resulting in a rougher interface (1 nm roughness as compared to 0.3 nm for Ar-sputtering) (60). Increasing the argon pressure (such that the vapor species are thermalized) mostly induces a coarser, granular microstructure with voids correlated with an increased coercivity but a reduced anisotropy of the multilayers. In the low energy processes, a superior [111]-texture is observed, which is probably due to less nucleation and a more rigorous nuclei selection. The recording results for Kr-sputtered MLs are similar to those for evaporated MLs.

K_u can also be increased by suitable dielectric underlayers that also may serve for optical enhancement, eg, ZnO and In_2O_3 (61). They improve the [111]-texture because the [002]-planes of these materials grow in parallel to the substrate surface; ZnO has in addition a small misfit (<2%) to Pt-Co. These layers increase, however, the noise of the disks, probably due to polarization fluctuations.

As in the case of RE-TM alloys, the properties relevant for recording are determined by proper material selection and by the preparation process. *Comparison of RE-TM films with Pt/Co Multilayers.* In Figure 12 the temperature dependence of the magnetization is schematically shown for a ferrimagnetic alloy (eg, amorphous RE-TM films) in comparison to a ferromagnetic metallic multilayer (39). The magnetic properties of Pt-Co multilayers are less favorable for MO recording than those of RE-TM films. First, they exhibit a higher Curie temperature (>300°C) implying less thermal sensitivity in recording which is only partly compensated by their smaller film thickness. Therefore, there are efforts under way to reduce the Curie temperature without reducing the Co content, eg, by adding Os or Re to Co (62). Second, they do not have a compensation temperature, leading to a high magnetization in the temperature range of domain formation. Due to demagnetizing effects, this favors subdomain formation during writing and stability of residual domains during erasing. Therefore, higher magnetic fields are needed for writing and erasing in comparison to RE-TM-based disks (63).

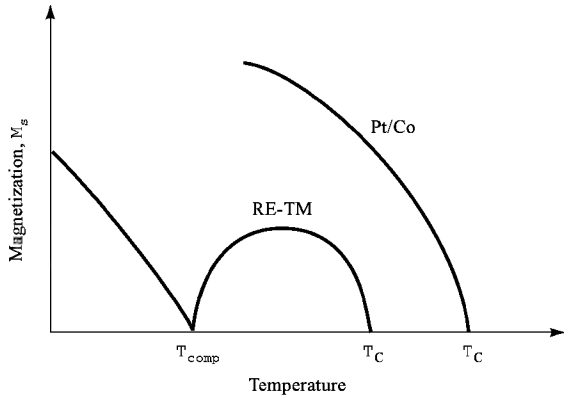


Fig. 12. Temperature dependences of the magnetization M , one curve typical for ferrimagnetic films, eg, RE-TM or garnets, the other one typical for ferromagnetic Co–Pt multilayers (39). T_{comp} – compensation temperature; T_C – Curie temperature.

An advantage of Pt–Co ML in future blue recording is their higher Kerr rotation in the visible wavelength region leading to a higher figure of merit (FOM). The purpose of blue recording is to get smaller domain sizes, but a higher FOM does not always lead to a higher signal-to-noise ratio (SNR). The signals and different noise levels expected for the two media classes are schematically depicted in Figure 13, for the case that the system noise is negligible.

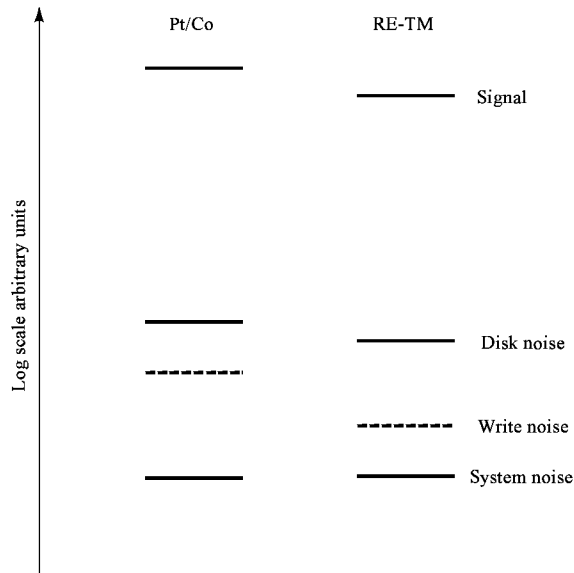


Fig. 13. Expected signal and noise levels for RE-TM alloys and Pt–Co multilayers (schematic). The total noise entering the SNR is the sum of the system noise, disk noise, and write noise. The system noise is electronic noise and photon shot noise and is comparable for disks with the same reflectivity.

If the predominant noise already stems from the unwritten disk and is not caused by magnetooptic effects, a higher signal proportional to $R \cdot \theta_K$, with R being the reflectivity of the layer, will indeed lead to a higher SNR. From the optical and magnetooptical constants a 3 to 6 dB higher FOM can be expected for Pt–Co. Results reported so far in the literature are related to this disk-noise limited range, and a few dB higher SNR is reported for Pt–Co in comparison to RE-TM. Examples are given below, where d/λ designates the length of the recorded domains and w/λ the laser wavelength used. The results for RE-TM disks obtained under the same recording conditions are given in parens.

SNR, dB, for ML	d/λ , μ	w/λ , nm	Reference
40 (38 dB for GdDyFeCo)	4.0	488	64
30 (24 dB for TbFeCo)	0.5	488	65
40 (37 dB for GdTbFe)	0.38	458	66
30 (20 dB for TbFeCo)	0.3	488	67

If the disk noise is suppressed by improved manufacturing procedures, the predominant noise is expected to be caused by irregularities of the domain boundaries (domain jitter), and signal and noise are equally enhanced by a higher FOM and the SNR is not improved. In this range (write-noise limited region) which extends for disks and ir recording up to about 58 dB, RE-TM based disks might have an advantage because of their finer microstructure. This is most important for domain edge recording.

If the microstructure becomes ever finer by improved deposition technology, the domain irregularities should diminish. The SNR is limited by the shot noise of the laser source and is equal to $R \cdot \theta_K^2$. In this region a high θ_K is of great value.

Magnetic Oxides. Mixed oxides of ternary iron and other metal ions that crystallize in a structure of spinel, garnet, or hexagonal type have also been investigated as MO media. These materials are discussed in References 68 and 69. The general composition of spinel is MFe_2O_4 where $M = Co, Mn, Ni, Zn$, and others (70). An example optimized for high Kerr rotation is $CoFe_2O_4$. Rare-earth iron garnets (RIG) have the formula $R_3Fe_5O_{12}$ where $R = Y, Gd, Tb, Dy$ (71,72). An example for a hexaferrite (73,74) is $Ba(CoTiFe)_{12}O_{19}$.

Promising candidates for MO recording among the ferrites are ferrimagnetic garnets of composition $Dy_{3-x}Bi_xFe_{5-y}(Ga, Al)_yO_{12}$ (71,72). The primary effect of Bi is a strong enhancement of the MO effect. Due to a strong intrinsic magnetostrictive effect, Dy enhances the perpendicular anisotropy, which is primarily stress-induced by the magnetostrictive effect.

These materials can be deposited by pyrolysis or by RF sputtering onto glass or garnet substrates. A deficiency of garnets and oxides is their high crystallization temperature ($>400^\circ C$) excluding polycarbonate (PC) as substrate material and making mass production difficult. Amorphous layers with garnet composition are paramagnetic at room temperature. The films become polycrystalline by *in situ* deposition on heated substrates or by post-annealing.

The garnets exhibit a high chemical stability (advantage over RE-TM), suitable magnetic properties, and high magnetooptical effects at shorter wavelengths (no clear advantage). In particular the presence of a compensation temperature allows tailoring of the magnetization and coercive field similar to RE-TM alloys (see Fig. 12), which is an advantage over Ba-ferrites and Pt–Co ML. Optimized disk structures of a 230-nm garnet film sputtered onto single-crystalline GdGa–garnet substrates have shown a SNR of 54 dB at 514 nm wavelength for a domain length of 1.4 μm (75).

The magnitude of H_c however, is determined by the stress-induced uniaxial anisotropy and by the grain size. The latter has to be small (<50 nm) to achieve high H_c but also noise levels must be sufficiently low. The polycrystalline nature of garnet films and thus the sensitive dependence of the grain size and its distribution on the deposition parameters is one of the crucial aspects of garnets for MO recording. Any irregularities in the morphology of the films lead to a drastic increase of the media noise reducing the SNR.

Exchange-Coupled RE-TM Layers. It is difficult to obtain RE-TM thin films that exhibit all the desired magnetooptic and micromagnetic properties. In most cases, the optimization of one property adversely affects another property. New and interesting possibilities in this regard utilize sandwich structures of two or more RE-TM films (exchange-coupled layers, ECLs) with different magnetic properties. One layer (the storage layer, s) can, eg, be optimized toward high coercivity for storage, and the other one (r) toward high Kerr rotation for read-out. Examples are $[GdFe(r) TbFe(s)]$ (73,74,76) or $[Nd_{10}Tb_{20}(FeCo)_{70} Nd_{5-10}TbFeCo]$ (77). For a trilayer stack $[200\text{ nm } Tb_{18}Fe_{49}Co_{33}(s) 10\text{ nm } Nd_{18}Co_{82}(r) 5\text{ nm } Tb_{18}Fe_{49}Co_{33}]$ (78), a high

coercivity of about 159 kA m (2 kOe) and $\theta_K = 0.37$ to 0.46° has been reported.

The most spectacular applications of ECLs are the possibility of direct overwrite (DOW) with laser modulation (79,80) and of magnetically induced superresolution (81,82). The stacks comprise at least a storage layer s and a bias layer b . For both applications, the storage layer s has the lower T_C and the higher H_c at room temperature when compared to the bias layer b . At room temperature, b is homogeneously magnetized (initialized) by an external permanent magnet (H_{initial} is about 400 kA m (5 kOe)).

DOW is achieved by switching the laser power between two levels. For erasing, the double layer is heated up to the Curie temperature of layer s , so that b remains unchanged and s is initialized locally during cooling due to an exchange coupling to b (avoiding a magnetic interface wall). New domains are written first into b by switching the laser to the higher level and copied to s during cooling.

When used for superresolution, the laser beam is incident on b , which hides the domains in s . During read-out, b is heated and the domains in s are copied to b . The optical system sees only the overlap area between the laser spot and the temperature profile which is lagging behind, so that the effective resolution is increased. Experimentally it is possible to double the linear read-out resolution, so that a four times higher area density of the domains can be achieved when the higher resolution is also exploited across the tracks. At a domain distance of 0.6 μm , corresponding to twice the optical cutoff frequency, a SNR of 42 dB has been reached (82).

Stability of Domains. In bilayers used for DOW or superresolution, the bias layer is homogeneously magnetized in the ground state after a complete write or read cycle, whereas the storage layer contains magnetic domains leading to magnetic interface walls between the two layers. For the domains to be stable, the wall energy density σ_w has to be lower than the coercive (area) density $2 \cdot H_c M_s \cdot d$ of both layers (thickness, d).

In order to achieve stable storage in exchange-coupled layers, different stack designs and material compositions have been used. By increasing the thickness of the bias layer, its coercive energy is proportionally increased; the thermal sensitivity of the stack, however, is decreased. As an example, for Tb–FeCo(s) TbDy–FeCo(b) a thickness of the bias layer b of 150 nm was reported (83). Instead of increasing the coercive energy, σ_w can be decreased by tuning the composition (84,85) properly or by intermediate layers.

Intermediate layers reducing the wall energy can be magnetic with low K_u or even possess an in-plane easy axis (86–88), so that the core of the wall is situated in this layer for a minimum energy configuration. This has been achieved with, eg, 3–6 nm Nd–FeCo (89) or with an interface nitride layer (1.5–2.2 nm Tb–FeCoN) (90). The intermediate layer can also be a nonmagnetic layer, softening the exchange between the two magnetic layers (eg, Pd) (91).

The trilayer configuration is a considerable improvement over the original concept of DOW with ECLs. A reduction of H_{initial} to 200 kA m (2.5 kOe) has been achieved because the magnetic walls, which have to be created during initiation, have a lower σ_w (92).

Another significant achievement has been obtained with a quadrilayer stack, where the initializing of b is taken over by the two additional layers, so that an initializing permanent external magnet is no longer necessary (93–95). The additional initializing layer i has a T_C above the writing temperature and carries a permanent magnetization. The switching layer is situated between i and (b/s). Its Curie temperature is below the writing temperature but above room temperature, such that i decouples magnetically from b during writing but initializes b during cooling.

Interface Walls and Domain Formation. The existence of a macroscopically large interface wall between two magnetic layers gives rise to new effects in the domain formation process that are not possible in single layers (96,97). The energy of the wall can be determined by reversal experiments (98,99). The main feature is a possible vertical motion of the interface wall across one layer, which can also be a mechanism to destroy domains in the storage layer of DOW-ECLs. It explains also the effect, where the coercivity of the intermediate layer in a trilayer stack depends on the relative orientation of the (TM subsystem) magnetization of the outer layers (100). For antiparallel orientation, an interface wall is always present and probably shifted across the intermediate layer. For parallel orientation, domains have to be nucleated for the high field branch of the hysteresis loop.

Layer Stacks and Protective Layers. The layer stack of an MO disk consists mainly of an MO layer, a dielectric antireflection layer, and a metallic reflection layer (Fig. 14). The thickness of the antireflection layer as well as that of the MO layer have to be properly chosen to obtain a maximum magneto-optical figure-of-merit (FOM). The FOM can be further increased by using a quadrilayer configuration with dielectric layers on both sides of the MO layer. Practical disks use the generalized configuration 50–120-nm dielectric layer, 25–90-nm MO layer, 17–70-nm dielectric layer (for quadrilayer configuration only), and 15–150-nm reflective layer.

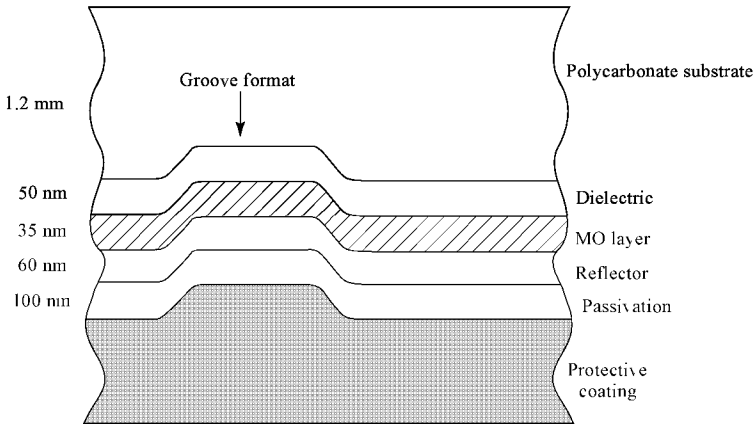


Fig. 14. Cross section of an MO disk (trilayer configuration) (101). Examples of the various layers are dielectric, AlN, Si₃N₄; MO layer, TbFeCo; reflector, Al; passivation, AlN.

The dielectric layers have to fulfill several functions. They must provide a barrier against oxygen and moisture, an antireflection layer for coupling in of the laser light, and heat insulation of the recording layer (in a quadrilayer configuration). Predominantly, AlN and amorphous Si₃N₄ are used as materials for dielectric layers (102), but Ta₂O₅, ZnS, and BaAlSiO sometimes are employed.

MO Media Summary. When compared to magnetic recording on hard disks, the advantage of MO data storage is the removability of the disks and the high storage capacity (especially on multiplatter (juke-box) systems) whereas the access times have not yet been reached.

The features of the discussed MO media classes relevant for recording are summarized in Table 3 (35,68). RE-TM based disks meet all requirements for the first generation of MO recording. Their main deficiency is the need of extra protection layers due to the lack of chemical stability. Pt–Co and garnet disks perform better in this respect, enabling a two-layer stack design. For high performance disks, however, dielectric and metallic layers are always necessary to optimize the optical and thermal properties of the disks. A disadvantage of Pt–Co is their higher magnetization, favoring subdomain formation and stable residual domains after erase, so that the magnetic field during writing and erasing has to be increased. Disadvantages of garnet disks are their high preparation temperature, excluding plastic substrates, and their coarse microstructure, causing noise. Which of these media classes is best suited for recording applications in the range of shorter wavelengths is not yet decided, but due to their fine microstructure and high versatility, RE-TM alloys are still good candidates.

Table 3. Properties of MO Materials

Material ^a	Microstructure	Deposition temperature, °C	H_c ^b kA/m	θ_K ^c	SNR ^d dB
α -GdTbFe	no grains	RT	~ 400	0.3	58
α -TbFeCo	no grains	RT	~ 400	0.3–0.4	61
Co-ferrite	0.1–0.5 μm	500	~ 250	10 μm	35 ^e
Ba-ferrite	40 nm	620	~ 100	1 μm	50 ^f
Co–Pt	10–30 nm	RT	~ 100	0.2	55

^a Chemical stability of α -GdTbFe and α -TbFeCo is poor; it is good for other materials.
^b H_c – coercive force. To convert kA/m to kOe, divide by 79.58.
^c θ_K = Kerr rotations in degrees.
^d Signal-to-noise ratio measured at conditions: wavelength (λ) = 800 nm, carrier frequency (f) = 1 MHz, linear velocity of the disk (v) = 5 m/s, bandwidth (BW) = 30 kHz, unless otherwise noted.
^e f = 0.5 MHz, v = 2 m/s, BW unknown.
^f f = 1.5 MHz, v = 2 m/s, BW unknown.

Phase Change Recording (PC-R). Erasable and rewritable optical data storage disks in MO technology exhibit in comparison to the widely available magnetic hard disks the possibility for fast disk exchange and in addition a broad insensitivity to dust and magnetic interference. Disadvantageous are their longer access time and a lower data writing rate. As explained earlier, the longer access time is caused by the relatively heavy and thus mechanically inert write–read heads of the MO technology, and the lower transfer rate for writing by the absence of a direct overwrite technique. Erasable and rewritable optical data storage disks in PC (phase change) technology unite the advantages of optical storage (high storage density, exchangeable media, robustness) with direct overwrite. On a PC medium the information is stored as amorphous spots in a thin crystalline film. The PC technology is based on fast, reversible transformations between crystalline and amorphous phases of certain alloys. Surveys on that topic can be found (103,104). Time and temperature plots of write, read, and erase processes in a thin PC layer are shown schematically in Figure 15 (105).

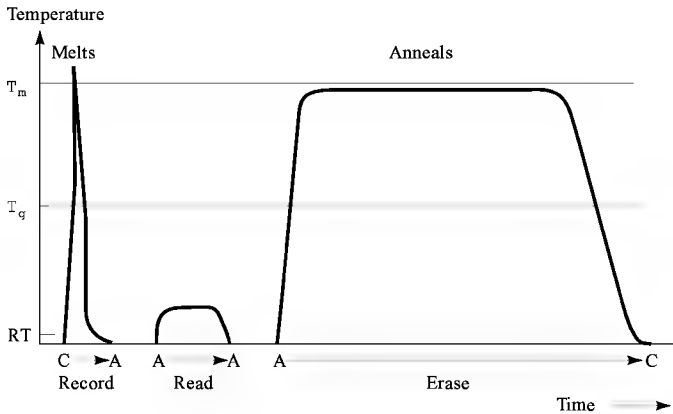


Fig. 15. Time–temperature transformation in a thin-phase change layer during recording–reading–erasing (3,105). C = Crystalline phase; A = amorphous phase; T_m = melting temperature; T_g = glass-transition temperature; RT = room temperature.

The writing process, that is, the transition crystalline → amorphous, is caused by briefly (<50–100 ns) heating up the selected storage area (diameter (ϕ) ca 0.5–1 μm) by a laser pulse to a temperature above the melting point of the memory layer (Fig. 15, Record), such that the film locally melts. When cooled faster than a critical quench rate (10^9 – 10^{10} K/s), the formation of crystalline nuclei is suppressed and the melted area solidifies into the amorphous (glass-like) state.

The reading of data is performed optically, based on the difference in reflectivity between the well-reflecting crystalline and the opaque and lower reflecting amorphous phase. A low power laser beam is used to avoid crystallization (Fig. 15, Read).

To erase information by the transition amorphous → crystalline, the amorphous phase of the selected area must be crystallized by annealing. This is effected by illumination with a low power laser beam (6–15 mW, compared to 15–50 mW for writing–melting), thus crystallizing the area. This crystallization temperature is above the glass-transition point, but below the melting point of the material concerned (Fig. 15, Erase).

Crystallization processes generally need longer times. The effective dwell time of a particular point of the disk under the laser spot, as determined by the data rate, is about 100 ns. To use this time for heating more effectively, it was proposed first to widen the laser beam elliptically in track direction from 1 to 1–15 μm thus gaining a longer annealing time for crystallization, although this solution would only have allowed erasure of information in blocks (106). Intensive research work has led to alloys that crystallize within about 50 ns when heated by a laser (103).

In PC technology, in contrast to MO technology, the writing is done as direct overwrite, this means that high data transfer rates in the region of 1–10 MByte/s can be reached provided the crystallization time is fast enough, similar to magnetic hard disks and twice as high as for nondirect overwrite MO storage. In addition, since no polarizing optics are needed, the optical read–write head can be built simpler and lighter, which means it is easier to move. This allows shorter position and access times for a PC system as compared to an MO system.

Special attention has to be given to the long-term stability of the information written in PC technology. Small amorphous areas in a crystalline matrix tend to recrystallize. For that reason, the intensity of the laser beam for reading is held very low, thus allowing a nearly unlimited number of read cycles. In comparison, a high number of write–erase cycles is critical. Besides the selection of suitable, long-term stable PC materials, sometimes the laser energy is adjusted to the number of cycles (107). Based on a crystallization temperature of 320°C and an activation energy of 3 eV, the life span of amorphous domains amounts to about 100 years (3).

Another advantage of PC technology is the planned and partially realized use of multifunction drives for operation of either CD-ROM, WORM, or PC-R disks alternatively in the same disk drive. This is technically feasible, since for reading (in CD-ROM, WORM, PC-R) as well as for writing (WORM, PC-R) similar principles and hardware are used (108). However, the reflectivity change of PC disks (40–70%) is in general lower than the CD-ROM standard (30–70%) requires.

Materials for PC Media. Crystalline alloys of elements from the fifth and sixth main group are preferred (3,103,109–111). As the first PC materials, tellurium suboxides as well as Te–Se or Te films that had been doped with small amounts of other elements like Ge, As, or Sb to shift the crystallization point to $>100^\circ\text{C}$ have been described.

Research has led to alloys which undergo laser-induced crystallization within about 50 ns. This is possible, for example, with TeGe alloys, which also possess the necessary temperature stability up to 180°C and exhibit sufficient reflection (crystalline phase) and transmission characteristics (amorphous phase), respectively. TeGe alloys have not found a practical use because of the formation of depressions in the memory layer typical for them after repeated

write-erase cycles (112). However, typical requirements on PC media and especially the stability-erasability trade-off can be illustrated with this system. Pure amorphous Te crystallizes very rapidly (good erasability) but is not stable at room temperature. Alloying a few atomic % of Ge increases the crystallization temperature (better stability) but also raises the minimum crystallization time above the practically available laser dwell time (about 100 ns). This holds equally true in the whole two-phase region, where crystallization implies a phase separation by long-range diffusion. The stoichiometric compound $\text{Te}_{0.5}\text{Ge}_{0.5}$, however, crystallizes by a polymorphic transformation without long-range diffusion and the necessary crystallization time falls below 100 ns.

More recent are investigations on ternary alloys from InSbTe, GeSbTe (113), TeSeGa, GeTeTi, and InSeSb; in some cases, even quaternary alloys are studied, such as TeGeSnAu. Detailed literature on this topic can be found in References 3,103,109, and 114. GeSbTe is the most studied. The key to achieving high speed is the design of an alloy where the crystallization process involves diffusionless crystal growth in a system that does not phase segregate. These materials can be crystallized with laser pulses of 30 ns duration and less.

GeSbTe alloys show excellent characteristics for such application, including fast transition times and excellent stability (115). They seem to be best suited for overwriting with linear velocities from 5.6 to 22 m/s. The pseudobinary phase diagram shows that suitable compositions lie on or close to the GeTe– Sb_2Te_3 tie line, eg, GeSb_2Te_4 , GeSb_4Te_7 , or the recently commercialized $(\text{GeTe})_2(\text{Sb}_2\text{Te}_3)_1\text{Sb}_{0.5}$ (116). The transition temperature (amorphous $\rightarrow$ crystalline) T_x increases with increasing GeTe content.

Of importance is the deposition technique for PC layers: suitable PC-materials are applied as thin layers (thickness 20 to 100 nm) predominantly via vapor deposition in vacuum onto proper substrate platters (glass, polycarbonate). To protect the memory-active PC layers from oxidation and contamination, they are embedded between two dielectric layers serving as diffusion barriers or heat barriers, respectively (SiO_2 , Si_3N_4 , AlN, ZnS). A typical layer stack consists, for example, of $[\text{ZnS} + \text{SiO}_2]$ mixture (180 nm) GeTeSb (25 nm) $[\text{ZnS} + \text{SiO}_2]$ (20 nm) Al (180 nm) (117). To improve the number of cycles, hard capping layers of MgF_2 , CaF_2 , or SiO_2 are preferably employed. As thermal barriers between the substrate and memory layer, intermediate layers of Si_3N_4 or AlN are utilized (101). The whole layer stack has to be optimized optically, thermally, and mechanically.

Projected Reversible Optical Recording. In the literature, a multitude of suggestions can be found for possible, reversibly operating optical recording techniques including the materials deemed suitable for this. Common to all these techniques is their failure to develop to commercial use.

Reversible Data Storage in Dye-Polymer Layers. In the 1970s, attempts were undertaken to reversibly store data in dye-in-polymer films. Corresponding studies were continued in the late 1980s, since pigment–polymer films promise advantages over MO-R and PC-R systems regarding specific performances, eg, signal-to-noise ratio of the reading signal and costs (layer application by spin coating). Three development directions have been pursued: (1) reversible generation of depressions, (2) reversible bubble generation, and (3) reversible deformation.

Reversible Generation of Depressions. This is done in highly dyed or pigmented polymer films by spotwise heating with a laser. Coloring materials, ie, soluble organic dyes or organic pigments with very low solubility in solvents and binders, and with high absorption in the near-ir, are used as recording media. The information can be erased by spotwise heating, where the high surface tension of the heated polymer film fills the depression (12,118). This development does not promise success, however, because of the time-intensive erase process (>100 ms).

Reversible Bubble Generation. In a two-layered pigmented polymeric layer system with ceramic coating, a bubble forms beneath the ceramic coating when heated locally (write process). For erasure, the bubble is closed again by heating the polymer close to its melting point (119). This development is also hampered by extensive erase times (>5 ms).

Reversible Deformation. This is done by spotwise laser illumination of a two-layered composition which consists of a dyed/pigmented thermoplastic and a dyed/pigmented viscoelastic polymer film: on writing, the lower thermoplastic polymer film is thermally expanded using the laser. The emerging deformation is fixed by the viscoelastic film. Erasure is accomplished by heating the thermoplastic film beyond its glass-transition temperature; the resetting forces of the upper viscoelastic layer reverse the deformation (120). With this system, interesting results have been achieved; the erase time is <500 ns, and $>10^3$ write-erase cycles have been reached. For practical use, however, the performance values of MOR or PCR layers should at least be reached approximately.

Materials with Reversible Coloring. A dye (or organic pigment) is called photochromic when it is rearranged into an energetically higher isomeric form together with a change of its absorption wavelength under influence of light of wavelength λ_1 (see CHROMOGENIC MATERIALS). The reverse reaction can be triggered by absorption of light of wavelength λ_2 , thermally, or by both processes simultaneously, where $\lambda_1 < \lambda_2$:

$$A(\lambda_1) \xrightleftharpoons[h\nu_2]{h\nu_1} A(\lambda_2) \text{ with } \nu_1 > \nu_2 \text{ and/or } kT$$

The difficulty of such a dye system is that only two variable characteristics (λ_1 and λ_2) have to fulfill three functions (writing, reading, and erasing). If the dye does not absorb, it cannot be switched; if it absorbs, it reacts (the so-called threshold problem). This predicament is circumvented in most cases by either using very low intensity λ_1 or λ_2 light, or using a third wavelength λ_3 for reading at which the dye does not absorb or react which increases hardware complexity.

Another impediment is the demand that both states should be stable in a thermodynamic sense, ie, of the same energy. Practically this cannot be realized; one of the two states will always lie on a slightly higher/lower energy level. For the whole system to remain at least metastable, a sufficiently high activation barrier has to lie between the two different energy levels, or the energetically higher level has to be supplied continuously with energy to avoid unwanted shifting of the switching states.

The advantages of dyes are their simple applicability to substrates (spincoating, sublimation, vapor deposition, etc) and their high signal-to-noise ratio on reading, assuming sufficiently high optical contrast (high molecular extinction). Figure 16 shows the maximal achievable SNR for reading data from different storage systems: nonionic squarylium dye (WORM system), GdTbFe-alloy (EOD on MO-R basis), and TeSeSb-alloy (EOD on PC-R basis) (121). The signal from PC and SQS, based on reflectivity changes, is very strong; it is about 10^3 times that of MO media based on RE-TM. In principle, the maximum SNR can therefore be higher for PC and SQS than for MO (Fig. 16). In practice, the obtainable SNR is, however, limited by noise from the written marks (eg, irregularities of their boundaries), so that at present the signal-to-noise ratios are comparable for the two media.

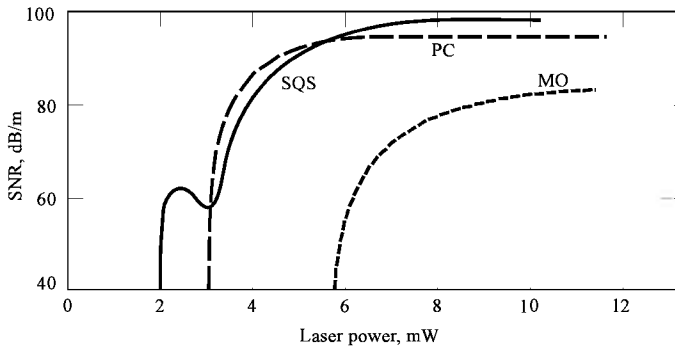
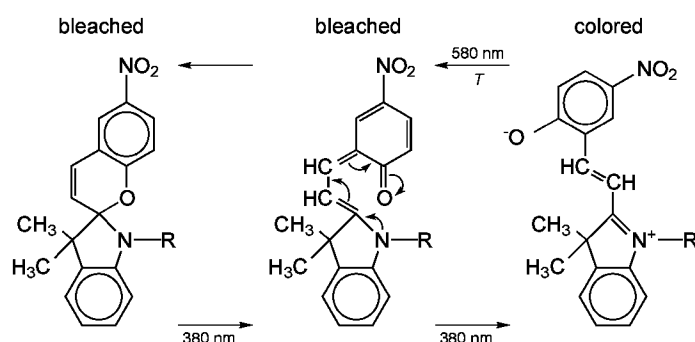


Fig. 16. Maximum achievable signal-to-noise ratio (SNR) on read-out of different writable optical data storage systems as a function of the writing energy (laser power) (121). SQS = Organic dye system (WORM); PC = phase change system (TeSeSb); MO = magneto-optical system (GdTbFe). See text.

Metallic Alloys with Optically Controllable Coloring. Besides photochromic organic dyes and pigments, metallic alloys with optically switchable coloring, dependent on crystal modification, were the subject of intensive research work in the 1980s. These consist of two or more metals like Al, Cu, Ag, or Zn (10). As an example, the alloy AgZn exhibits a hexagonal crystal structure and silvery color at room temperature. When heated briefly with a laser impulse to 300°C, after cooling down it forms a cubic phase with pink color. Heating to temperatures <300°C restores the initial structure and color (122). A practical application of metallic alloys with optically thermally controllable coloring, however, is not known.

Photochromic Organic Dyes. Intensive investigations into this category of substances have led to numerous patent applications. Copper-phthalocyanine pigments, organic dyes based on cyanine (Ricoh, Pioneer), naphthochinone (Nippon Denki), and benzothioapyrane (Sony) (123) have been described. They did not lead, however, to any commercial use. Surveys on the possibilities of optical data storage with photochromic dyes can be found (124,125).

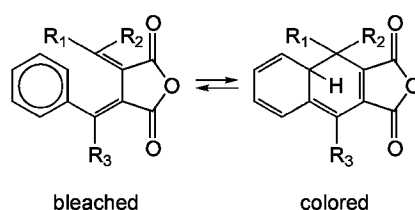
The dyes investigated can be divided into two groups based on the molecular mechanism of their switching. The first types work by electrocyclic opening and cis-trans-isomeric rearrangement. Spiropyrans (126) have found particular interest because of their intensive coloring, which allows application in thin layers and thus fulfills a prerequisite for small pit size and therefore high storage density. Problems are caused, however, by the spiropyran's lack of chemical stability. A typical representative of this category is the benzospiropyran (127) shown in the following.



From the colorless state it can be switched with light of short wavelength ($\lambda = 380$ nm) via an electrocyclic ring opening and cis-trans rotation of one half of the molecule into a state with violet-purple color. The reverse reaction is effected by visible light ($\lambda = 580$ nm). Since the system is metastable, one of the two reaction directions is matched by a rival thermal reaction, the thermoreversion. This progresses, however, in the case of benzospiropyran, at room temperature by a factor of 10^3 slower than the light-induced reaction.

Even though in a solution, all cis-trans isomeric dyes switch fast and almost effortlessly, show a drastic increase in switching time after embedding into polymeric media, and fatigue leading up to a standstill of switching after a few switching cycles. This increase in switching time and debilitation leading to immobility is caused by a steric impediment of free rotation of the dyes by the polymeric matrix (128,129). For example, benzospiropyrans can be switched in solution in terms of microseconds. After embedding into a polycarbonate matrix, the switching time spontaneously increases by several orders of magnitude. When switched repeatedly, the dye molecule reacts more and more slowly, until it is rendered completely motionless after about 10 switching cycles. This steric impediment is a consequence of relaxation processes of the polymer molecules at and after each switching, since the switching event demands a certain free volume in the polymer matrix for the rotation of the dye molecule. These relaxation processes lead to complete freezing of all motions; that is, a standstill of the switching process in the dye molecule.

The second type of mechanism offers reversible switching without space-consuming transfer processes. These dyes are electrocyclically reacting dyes, where the transition from one form to another form happens by electron shift or proton transfer, without the rearrangement of bulky molecule segments. Fulgides have been proposed for this purpose. Fulgides are bis(methylene) amber acid anhydrides (130), eg, (*E*)-3[1-(2,5-dimethyl-3-furanyl)ethylidene]-(4-isopropylidene-2,5-furandione [59000-86-1], $C_{15}H_{14}O_6$, where $R_1 = R_2 = R_3 = CH_3$ and the phenyl ring is replaced by a 2,5-dimethyl furan ring (131). The combination of fulgides with polymers, especially with LC polymers, has led to photochromic recording media for holography and to nondestructive optical read-out systems (132). Dyes from the material class of the fulgides have the disadvantage of generally small molecular extinction (optical contrast differences are too low).



New photochromic dyes with electrocyclic reactions have been proposed on the basis of 1,5-electrocyclization of heterogenous pentadienyl-anions (124). Still newer are investigations into the photocyclization of 2,4,6-tri-isopropylbenzophenones for vinyl polymers in the glassy state (133).

The practical use of photochromic dyes as memory layers in erasable and rewritable data storage disks fails not only because of their physical limitations (lacking sensitivity, insufficient stability, low number of cycles), but also because the diode lasers required for switching in the visible range (wavelength between 450 and 600 nm) and the uv-range (around 350 nm) are not available.

LC Side-Chain Polymers with Dyes. All photochromic bistable dyes mentioned have the basic disadvantage that their long wavelength absorbing form is thermally unstable even at room temperature. For that reason, the molecular rearrangements effected by the photoreaction (including the color changes caused by these) are not permanent. To improve long-term stability, storage devices have been developed in which the backbone forms a liquid crystal polymer with mesogens as side chains (LC side-chain polymer). These polymers contain dyes of different nature which are responsible for the switching process. The composition of an LC side-chain copolymer with embedded dyes is schematically displayed in Figure 17 (134). These dyes are, for example, azobenzenes, spiropyrans, or fulgides and the polymer systems are LC polyacrylates and LC polyesters (135), LC polysiloxanes with spiropyran side groups (132,136), and LC polymers based on methacrylate copolymers with fulgimide side groups (137).

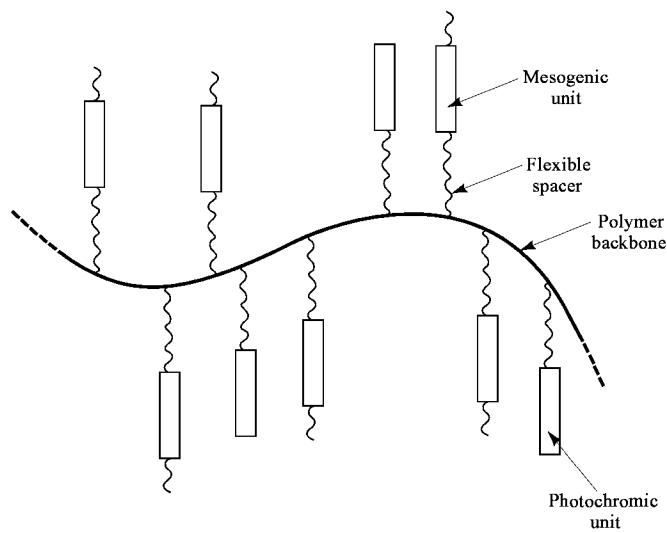


Fig. 17. Schematic structure of an LC side-chain polymer with embedded dyes (134).

In the simplest case the dye-induced storage process is based on a thermally induced phase change between disorder (isotropic melt) and nematic, smectic, or cholesteric ordering of the polymer molecules. Combined with the phase change are corresponding changes in the optical properties. The single steps of a write read erase cycle for an LC side-chain polymer with side chains containing dyes is outlined in Figure 18 (134).

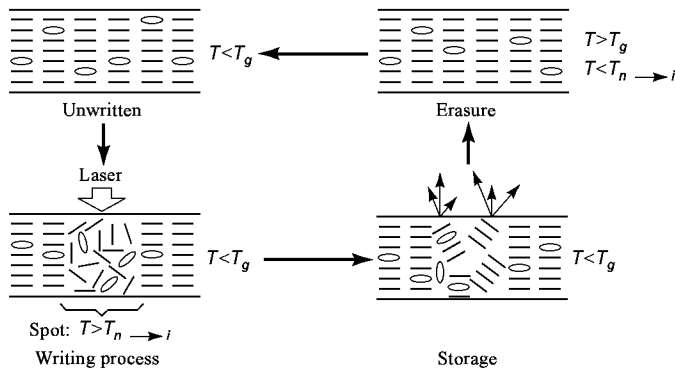


Fig. 18. Single steps of a write-read-erase cycle for an LC side-chain polymer (134). (—), mesogenic group; (O), dyestuff. $T_n \rightarrow i$, transition temperature nematic $\rightarrow$ isotropic.

In this case, the dye functions only to convert light energy into thermal energy. This process, thermorecording (137,138), has no technical use due to its large spot diameters caused by the heat transfer of the illuminated spots and its long switching times. In other proposals the phase change is triggered by the spotwise heating of a heat absorbing (generally metallic) substrate of the LC film.

If a dye is used which changes its properties by illumination, eg, its shape like the azobenzene, the phase change temperature can be lowered. The writing process then is nearly isothermal which helps to reduce spot diameters.

This photochemical influence on the shape of systems containing azobenzenes has been investigated. The optically induced transition of azobenzene from the trans into the cis state corresponds to a transition from a mesogenic (stretched) to a nonmesogenic (angled) shape. This causes a disturbance of the liquid crystalline order and leads to a change in birefringence (135). In extreme cases, the order can be destroyed completely; the process then is similar to the thermorecording process depicted in Figure 18.

The photochemically induced phase shift has been investigated (139). The best results were reached when the dye was not mixed physically, but copolymerized into the polymer (140). A big advantage of LC side-chain polymers is the ability to store information at room temperature and thus in the glassy (solid) state, so the information is conserved even when the photochromic dye isomerizes back thermally. Another advantage regarding long-time stability is gained by using photochromic dyes with thermally irreversible photochemistry. This is the case when using special optically reversibly switchable fulgides as mesogenic side groups.

Especially favorable storage effects are achieved by using linearly polarized light instead of unpolarized or circular polarized light. The optical axis of LC polymers in this case is reoriented by up to 90° relative to the polymerization plane of the light beam by cis-trans-isomerization cycles of azobenzene (141). The reorientation process occurs in the solid glassy state and leads to spots with stable phases in the polymer. The resolution is only limited by the wavelength of the writing beam. The induced modulations of structure and optical properties can be erased by global or local heating above the glass temperature. This process has been used for optical storage in homeotropic oriented polymer films (142) and for optical storage in cholesteric systems (143).

The mentioned polymers are in principle suitable for digital data storage, but the practical storage systems introduced up to now are based on polarization holograms. In these systems, use of linearly polarized light creates birefringence patterns which contain information on the polarization state of the light used for writing (144). It is not only possible to store information on phase and amplitude holographically, as in the common holographic storage media, but also in the polarization state (145) (see HOLOGRAPHY).

In Reference 146 three methods for playback of data are presented, which are superior to the usual method based on the contrast in light scattering: readout of dichroism, perpendicular readout mode, and readout of birefringence. Some critical comments also are given about global or selective erasure of data.

Different research groups have examined the special characteristics of LC polymers with photochromic dyes when going to thin films produced by spin coating (147) and ultrathin films produced by the Langmuir-Blodgett (LB) technique (148). Another development variant couples a command surface coated with photochromic dyes with adsorbed, nematically oriented LC polymers (149), sometimes using ultrathin films produced by the LB technique (150). In Figure 19 this principle is demonstrated by a command surface using an LB film of an amphiphilic azobenzene side-chain polymer with a backbone of poly(vinyl alcohol) (151,152). A photoinduced cis-trans rearrangement of azobenzenes induces a reversible rearrangement of a thick nematic LC polymer attached to the command surface (153).

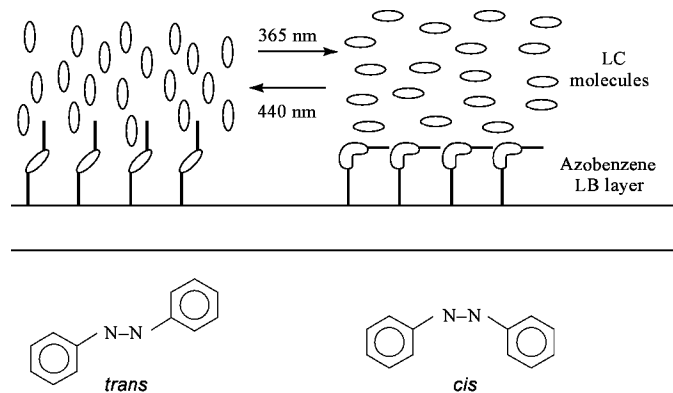


Fig. 19. Reversible data storage by means of a command surface (149,151–153).

The application of nonlinear optical recording techniques for reversible optical data storage based on the excitation of photochromic molecules by two-photon processes also has been described (154).

Finally, the use of photoreversible change of the circular dichroism for optical data storage is of interest. This technique offers an advantage over photochromic materials in that the data can be read in a way that does not damage the stored information. These chiroptic data storage devices have been demonstrated with the example of chiral peptides with azobenzene side groups (155).

All LC polymers mentioned herein generally are suitable for digital data storage, especially holographic data storage, because they are thermally stable at room temperature. A disadvantage, however, is the long switching time which is in the order of milliseconds to seconds. For that reason, these polymers currently are not acceptable for fast optical data storage.

Biopolymers for Reversible Data Storage. An interesting development for reversible data storage using photochromic materials employs bacteriorhodopsins (BR) optimized by genetic engineering (156,157). BR is the key protein in the photosynthesis system of the bacterium *Halobacterium halobium*. This bacterium can survive in extreme environments, eg, in concentrated salt solutions or very low temperatures. BR is embedded as a two-dimensional crystalline lattice into the lipid layer of the cell membrane, the so-called purple membrane (PM). A photochrome unit is linked to the polypeptide chain of the BR.

Under influence of light, BR acts as a proton pump to generate adenosine triphosphate (ATP). The multistaged photoprocess is depicted in Figure 20 (158). When illuminated with light of 520 nm wavelength, the photochemical reaction B to J is induced; the photochemical transition M to B is triggered by photons of 412 nm wavelength. These photoreactions lead to local changes in optical absorption and index of refractive in the PM film. To read and erase data, preferably in the form of holograms, both photoreactions can be employed: B-type holograms ($B \rightarrow M$) and M-type holograms ($M \rightarrow B$).

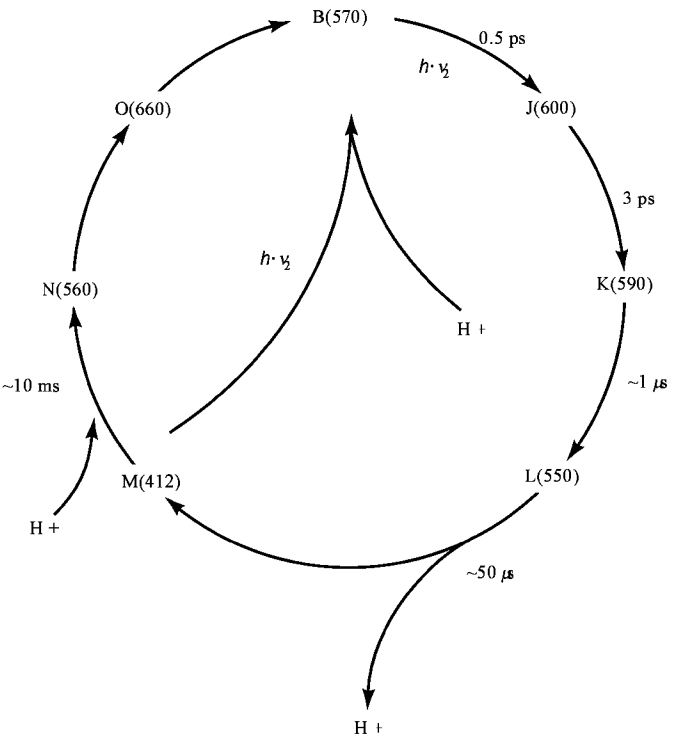


Fig. 20. The photocycle of bacteriorhodopsin with its intermediates including lifetimes and absorption wavelengths in nm in parentheses (158).

A technical use of these processes is possible and of high interest because of the special advantages of this system: high quantum yield (0.7), high recording sensitivity ($10^2 - 10^3 \mu J \text{ cm}^{-2}$), good line resolution ($>5000 \text{ lines mm}^{-1}$), and high number of write-erase cycles ($>10^6$). The shelf life of the written information, however, must be substantially improved. The biological material (so-called wild type) has a duration of the M state of only 10 ms due to thermal relaxation; in contrast, with genetically engineered mutants, durations of over four minutes have been achieved (159,160). Further development concentrates on breeding of BR mutants with a permanent M state at room temperature.

The purple membrane is harvested semiindustrially from halobacteria mutants which are bred in fermenters. The BR is then embedded into a polymeric matrix of poly(vinyl alcohol) or polyacrylamide. The BR films manufactured in this way are used for different applications, preferably in holography, for example, as a reversible transient data storage system for optical information processing (159). Another example is real-time interferometry by using the property of BR films to integrate over time (160). BR has been proposed also as a two-photon memory material because of its unusually large two-photon cross section.

Permanent data storage in BR films at room temperature is possible by adding hydroxylamine. In the illuminated areas of the BR film, retinaloxime is irreversibly generated which absorbs at 360 nm. For nondestructive reading, light in the near ir (750 nm) is used (158).

Compared to photographic films, BR is handicapped due to its very low light sensitivity. Also, as a possible material for disk-shaped digital optical information storage, BR does not offer an advantage over established magneto-optical materials in spite of its higher line resolution, because in both cases

maximum resolution is determined by the limited wavelength of laser diode light.

CARD-SHAPED OPTICAL DATA STORAGE

All commercial magnetic mass storage devices (floppy, hard disk, tape) as well as the optical mass storage devices (CD-ROM, WORM, EOD) require mechanical drives for writing and reading data. Mechanical drives, however, are cost-intensive, bulky, and noisy. They require positioning and access times in the order of milliseconds to seconds, which impedes fast writing and reading of data. Finally, they consume high levels of power, limiting the usable time of portable computers running on batteries.

A possible way of avoiding these disadvantages is offered by solid-state disks, Li-battery-buffered S-RAMs, and EEPROMs. Of special interest are the newly developed flash memories (flash-EPROMs). The last, however, are not freely removable; in addition they are rather expensive and their capacity is limited to around 20 MByte (161). A solution, especially as an alternative to CD-ROM and WORM disks, with the advantage of freely removable media, is offered by optoelectronic memory cards, which can be operated without a mechanical drive. Rather than storing exclusively alphanumeric data like the current magnetic devices or chip-cards, optoelectronic memory cards offer the possibility of storing pictures, graphics, drafts, fingerprints, possibly even sounds, as well as alphanumeric data. Such a credit card-sized storage medium is called optical memory card (OMC).

The composition of an optoelectronic memory card (eg, Laser Card of Drexler Technology Corp.) (162) is outlined in Figure 21 (163). Primary elements are polycarbonate foils with thicknesses of 250 to 400 μm, respectively, that are employed because of their high operating temperature and their good mechanical, optical, and dielectric characteristics. The OMC can be used as a ROM or a WORM media. Both possibilities of information deposition can be used separately or in combination.

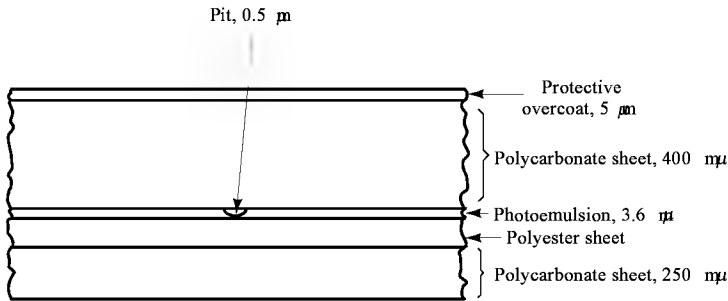


Fig. 21. Principles of optical memory card construction (Laser Card) (163).

An example in the medical field is use as a patient card. Basic data (name, address, date of birth, etc) are written as ROM, but the results of ongoing examinations are written as WORM by the doctor in attendance. For data security, code- and passwords can be used.

One of the benefits of an OMC is its immunity against static electricity and magnetic fields. Its capacity is 4.11 MByte in the version mentioned when used as a ROM, 2.6 MByte for the WORM version.

Data Deposition in Three and More Dimensions

HOLOGRAPHIC INFORMATION STORAGE

The techniques described for 2-D storage media use monochromatic laser light for writing and reading. The finite diameter of the laser beam and the optical limit defines the maximum planar storage density using light in the near ir to 10⁸ bit cm². Data storage technology aims to concentrate data in the smallest possible space, that is, to achieve maximum spatial storage density. This can be realized by three-dimensional data deposition as a bit oriented format or as an interference pattern (hologram). Therefore, holographic storage techniques attract great interest (164).

The highest information density is achieved by storing information in superimposed volume-phase holograms, so-called thick holograms. The thickness of the storage material must be greater than the wavelength by a factor of >100, so that several thousand holograms can be stored by angle-selective superimposition in the same volume element. With each new hologram the incidence angle is altered by a fraction of a degree.

In a setup for parallel recording and read-out processes in transient holography (158) the superposition of the information-carrying recording beam with a reference beam generates an interference pattern in the recording material, which leads to a spatial distribution of the index of refraction. The information can be read out observing the Bragg condition (see HOLOGRAPHY).

Of special interest is 3-D optical data storage by two-photon excitation (165), using two orthogonal laser beams having different wavelengths (166,167) or, in a simpler scheme, using a single highly focused monochromatic beam (168). Single-beam addressing potentially offers access to extremely high capacity multilayered disk-type storage media.

Materials. For holographic information storage, materials are required which alter their index of refraction locally by spotwise illumination with light. Suitable are photorefractive inorganic crystals, eg, LiNbO₃, BaTiO₃, LiTaO₃, and Bi₁₂SiO₂₀. Also suitable are photorefractive ferroelectric polymers like poly(vinylidene fluoride-*co*-trifluoroethylene) (PVDF TFE). Preferably transparent polymers are used which contain approximately 10% of monomeric material (so-called photopolymers, photothermoplasts). These polymers additionally contain different initiators, photoinitiators, and photosensitizers.

When exposed to light, the monomeric material in the photopolymers or photothermoplasts polymerizes, thus locally increasing density and index of refraction. A subsequent fixation process polymerizes the monomer throughout the polymer matrix.

Examples of photothermoplasts include polyacrylates, polyacrylamides, polystyrenes, polycarbonates, and their copolymers (169). An especially well-researched photothermoplast is poly(methyl methacrylate) (PMMA), which is blended with methyl methacrylate (MMA) or styrene as a monomer, and titanium-bis(cyclopentadienyl) as a photoinitiator (170).

Using suitable polymers, the analogously recorded information can be erased blockwise by illumination with light of appropriate wavelength. As an eligible photosensitizer, 2,2-dimethoxy-2-phenylacetophenone has been described (171).

Besides the classical photothermoplasts, LC side-chain polymers with distinct phase changes also are well suited for holographic purposes, and biopolymers from genetically engineered bacteriorhodopsine (BR) have been discussed as a holographic material.

For two-photon memories, a number of media types and reading mechanisms have been used (165). Generally, media comprise two photon-absorbing chromophores dissolved within a solid polymer matrix. Suitable reversible photochromic dyes are, for example, spiropyrans. Although photochromic materials often suffer from photobleaching, as well as from instability leading to self-erasure, new materials and host environments are under development (172). Bacteriorhodopsin (BR) also has been proposed as a two-photon memory material.

All attempts to develop photopolymers or photothermoplasts suitable for fast and reversible recording and read-out of volume-phase holograms, however, have not gained commercial application. The most important characteristics of materials for holographic information recording are listed in Table 4 (158).

Table 4. Materials for Holographic Information Storage^a

Materials	Spectral range, nm	Resolution, lines /mm	Exposure required, μJ /cm ²	Diffraction efficiency, ^b %	Reusability ^c
silver halide film ^d	400–700	1.000–10.000	10 ⁻¹ –10 ²	5–60	
dichromate gelatin ^d	uv–500	< 3.000	10 ⁴	30	
photoresists	uv–650	200–1.000	10 ³ –10	10–85	
photopolymeric materials	300–500	5.000–10.000	10 ² –10 ⁴	50	± 10 ³
photothermoplastic materials ^d	400–650	500–1.200	10 ¹ –10 ²	10	± 10 ⁴
magneto-optic materials	700–ir	> 1.000	10 ⁴ –10 ⁵	< 0.01	±
photorefractive materials					
LiNbO ₃	350–500	1.500	10	20	±
Bi ₁₂ SiO ₂₀	350–550	10.000	10 ³	25	±
photochromic materials					
inorganic	300–700	5.000	10 ⁴ –10 ⁵	1–2	± 10 ³
organic	400–800	5.000	10 ² –10 ⁴	1–2	± 10 ³
biological ^e	400–700	> 5.000	10 ² –10 ⁵	3	± 10 ⁶

^a Ref. 158.
^b Ratio diffracted / absorbed: irradiated intensity.
^c Approximate number of times to reuse.
^d Developing necessary.
^e Purple membrane from bacteriorhodopsin.

PHOTOCHEMICAL HOLE BURNING

Aside from holography, photon-gated spectral hole burning (photochemical hole burning (PHB), also named persistent spectral hole burning (PSHB)) is another possibility for achieving multidimensional information storage (173). Besides the three spatial dimensions, the parameters of frequency and electrical field strength can be used to store information.

If dye molecules are embedded into an amorphous matrix, preferably transparent polymers, greatly and inhomogenously broadened spectral lines are observed. This broadening is caused by the energetic interaction of the dye molecules with the locally different environment in the polymer matrix. The ratio of the homogenous initial line width of the dye molecule Γ_h to the inhomogenous line width of the dye in the polymer Γ_i ranges from 1:10³ to 1:10⁴.

If the dye in the polymer is illuminated with frequency stabilized laser light, molecules with absorption frequency identical to the frequency of the laser light are excited. After excitation, the absorption spectrum of the dye-doped polymer exhibits a hole at the location of the laser frequency with the width of the homogenous line of the dye (Fig. 22) (1,173). This process can be used for building high density data storage systems. The storage density, which is limited in planar data storage to <10⁸ bit /cm² by the frequency of laser light, is increased by frequency selective recording and reading, for example, at $\Gamma_i : \Gamma_h = 10^3 : 1$, to 10¹¹ bit /cm².

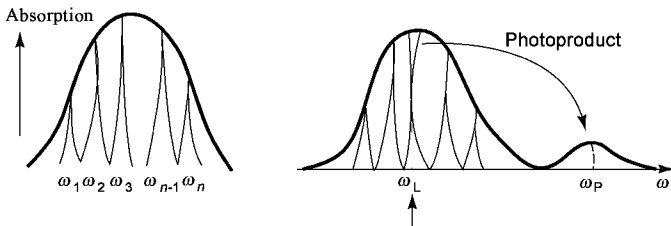


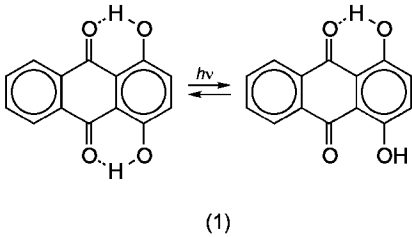
Fig. 22. Photochemical hole burning (PHB) (1,173) where ω is frequency; ω_L , frequency of the laser; and ω_P , frequency of the photoproduct.

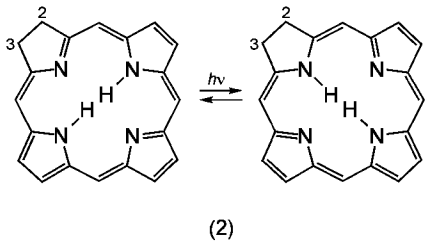
Another possible advantage of the PHB method is its potential multiplexing characteristics; by synchronous recording and read-out, exceptional data transfer rates can be achieved. These high transfer rates are of equal importance in practical data technology as high storage densities.

An interesting development of the PHB technique leads to four-dimensional data storage. By variation of an electric field applied to the sample the spectral profile of the absorption holes can specifically be altered. This adds two more dimensions to the geometrically two-dimensional matrix: frequency of laser light and electrical field strength (174).

A great disadvantage of PHB is the necessity to operate at very low temperatures (<20 K). Therefore, this recording technique currently has no practical significance; but it is subject to intensive research activity (175). One future aspect which may be important, if room temperature materials become available, is the usage of inexpensive semiconductor lasers in the near ir-regime (176).

Materials. Beside inorganic materials (eg, barium chloride fluoride crystals, doped with 0.05% samarium), transparent thermoplasts are preferred for the PHB technique, eg, poly(methyl methacrylate) (PMMA), polycarbonate, and polybutyral doped with small amounts of suitable organic dyes, organic pigments like phthalocyanines, 9-aminoacridine, 1,4-dihydroxyanthraquinone [81-64-1] (quinizarin) (1), and 2,3-dihydroporphyrin (chlorin) (2).





Current research aims at high efficiency PHB materials with both the high speed recording and high recording density that are required for future memory applications. To achieve this aim, donor–acceptor electron transfer (DA-ET) as the hole formation reaction is adopted (177). Novel PHB materials have been developed in which spectral holes can be burnt on sub- or nanosecond time scales in some D-A combinations (178). The type of hole formation can be controlled and changed between the one-photon type and the photon-gated two-photon type (179).

New PHB materials are composed of Zn-tetrabenzoporphyrin–aromatic cyanide–poly(methyl methacrylate) (180) or of tetraphenylporphyrin derivatives dispersed in polymer matrices such as PMMA and polyethylene (181). A survey of such materials has been given (181).

Substrate Materials for Optical Memories

High demands are placed on the substrate material of disk-shaped optical data storage devices regarding the optical, physical, chemical, mechanical, and thermal properties. In addition to these physical parameters, they have to meet special requirements regarding optical purity of the material, processing characteristics, and especially in mass production, economic characteristics (costs, processing). The question of recyclability must also be tackled.

The birefringence of substrate materials for optical data storage devices requires special attention, especially in the case of EOD(MOR) disks. Birefringence has no importance for glass substrates (glass does not exhibit any significant birefringence) and is only a subordinate factor for polymeric protective layers of aluminum substrates because of their reflective read write technique.

In the case of polymeric substrate materials, high birefringence not only makes exact focusing of the laser beam for reading and writing more difficult, but also effects an unwanted rotation of the plane of polarization during reading, which can severely impair the signal-to-noise ratio, especially using EOD(MOR) disks. In addition, unwanted oscillations in laser intensity are amplified by the reflection of a part of the elliptically polarized light generated by the substrate back into the laser. This feedback, called laser feedback power (LFP), depends on the magnitude of birefringence and the numeric aperture (NA) of the object lens. In a polymeric substrate disk, birefringence is caused by internal stresses in the material (stress birefringence) and or by preferred orientation of polymer molecules with anisotropic optical polarizability (orientation or rheo-optical birefringence). Therefore, polymers with low anisotropy of their molecular optical polarizability are preferred.

The anisotropy of polarizability can be positive (eg, polycarbonate) as well as negative (eg, polystyrene). This offers the possibility of minimizing birefringence by copolymerization or blending of suitable polymers with the right mixture ratio, eg, blends of poly(phenylene ether) (PPE) and polystyrene (PS). The magnitude of birefringence of axial-symmetrically oriented polymers vs their molecule orientation has been described (182).

Lower birefringence can be achieved by concerted efforts during injection molding, meaning an optimized tuning of the temperatures of material and tool injection rate, programming of an injection compression sequence, dynamic pressure, holding pressure, etc (163,183–185). This care can avoid frozen stresses as well as a too anisotropic orientation of the polymer molecules. Furthermore, polymers exhibiting a low anisotropy of their molecular optical polarizability are preferably employed.

To efficiently drive the development of improved substrate materials, the limiting values of birefringence have to be known; this is especially true for WORM and EOD(MOR) substrate disks. These limit values were laid down by the ANSI (American National Standard Institute) Technical Standard Committee (186–188). For 5.25 in. WORM disks, the ANSI document X 3 B 11 88-144 recommends a maximum LFP value of 9°; this corresponds to an optical path difference perpendicular to the plane of the disk of not more than 80 nm/mm (double path). For 5.25 in. EOD(MOR) disks, more stringent conditions apply (ANSI-document X 3 B 11 88-049), which also allow calculation of the allowed range.

Depending on the type of memory, further and or increased demands are imposed. For CD-ROM precise forming of information-containing pits (width 0.6 μm, depth 0.12 μm, length 0.8–3.3 μm) is necessary as are low melt viscosity and easy mold release for high production yields (equals short cycle time). For WORM and EOD(PCR), increased resistance to heat softening (heat distortion temperature (HDT A)>150°C) is needed. For EOD(MOR), very low birefringence (Δn < 20 nm/mm) is required and negligible warp from centrifugal forces at high rotation speeds (tilt angle < 4 mrad at 1.800–3.600 rpm, respectively). There must also be minimal residues of ionic halogen and mold release agents.

With disk diameters above 5.25 in., all parameters, eg, water absorption and thermal expansion, become more critical which aggravates the expansion or warp of disks. If in the future disk rotation speeds have to be increased significantly to boost data transfer rates, higher demands will be placed on warp (tilt angle) and modulus to avoid creeping (ie, irreversible elongation in radial direction). A survey of the requirement profile for the substrate material of optical disks is given in Table 5 (182,186,187,189).

Table 5. Requirements on Substrate Materials for Optical Memories

Properties	Remarks	Nominal value
<i>Optical properties</i>		
light transmission, %	820 nm	> 90
birefringence, nm/mm	single path	< 40(< 20) ^a
optical purity	θ > 200 μm	no particles
<i>Mechanical properties</i>		
thickness, mm		1.20 ± 0.05
thickness, tolerance, μm		±50
flatness (tilt), degree		±0.6
surface roughness, nm		< 15
tensile stress, MPa ^b	at yield	> 60
impact strength, kJ/m ^{2c}	Izod	no failure
<i>Thermal properties</i>		
vicat softening temp, °C	VST B 50	> 120(> 170) ^d
heat defl temp, HDT A, °C	(1.80 N/mm ²) ^b	> 110(> 150) ^d
<i>Physical chemical properties</i>		
water absorption, %	immersion	< 0.35
water absorption, %	23°C, 50% RH	< 0.15
<i>Processing properties</i>		
melt flow index (MFI), g/10 min	300°C 1.2 kg	> 60
cycle time, s		< 7

^a For EOD(MOR).

^b To convert MPa (N mm²) to psi, multiply by 145.
^c To convert kJ m² to ftlbf in², divide by 2.10.
^d For WORM and EOD(PCR).

NONPOLYMERIC SUBSTRATE MATERIALS

MO disks have to meet particularly high demands in terms of low birefringence; for WORM and EOD(PCR) disks a higher resistance to heat softening is wanted and for all optical storage disks with diameters exceeding 5.25 in. (>130 mm), increased requirements exist regarding smallest warp (thermally or by moisture absorption), distortion, and creep (at very high rotation speeds). These increased demands can be met partially by uv-curable cross-linked polymers, but especially by glass.

Glass. The distinct advantages of glass over thermoplastic polymer materials are high surface quality (surface roughness <3 nm); insensitivity to very high as well as very low temperatures without detrimental expansion contraction and without warp; no water absorption (water can lead to erosion of sensitive magneto-optical materials and phase change materials, and of the protective layer also); insensitivity to aggressive chemical vapors and liquids; impermeable to gases and vapors; and extremely long shelf life (>25 yrs).

Among the acute disadvantages of glass substrates are high weight; fragility, ie, extremely low fracture toughness (except for special armor glass); and high cost due to intricate, expensive manufacture and the need to apply the formatting layer in a photopolymerization (2P) process in a second, costly production step.

Glass substrates are used commercially for 5.25 in. EOD(MOR) disks only by a few manufacturers (eg, Philips, Hitachi); in contrast, for CD-ROM and WORM memories with disk diameters exceeding 5.25 in. (eg, 12 in. and 14 in.), glass substrates are employed frequently.

Aluminum. Some manufacturers also have WORM disks above 5.25 in. on offer with aluminum as substrate material. For Al the same advantages apply as for glass with the exception of a high coefficient of thermal expansion and lacking resistance to aggressive chemical vapors and liquids.

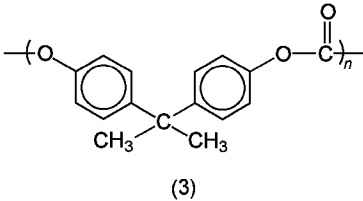
An advantage of aluminum is the high level of knowledge and the automated production plants stemming from the mass production of Al substrates for magnetic hard disks; these can be widely used for the production of substrate disks for optical data storage.

The limitation to disk constructions with a laser beam reflected at the disk surface is a large drawback, however. This prevents the insensitivity against dust and dirt, which is well known from current optical storage devices with a laser beam reflected after penetration of the transparent substrate. The distance between the disk surface facing the optics and the memory layer naturally has to be much smaller than in common optical disks, where the memory layer is deposited behind a 1.2-mm thick transparent glass or polymer substrate disk.

Another disadvantage of Al substrates is their higher weight compared to polymer substrates, and particularly the significantly higher production costs of the disks themselves.

POLYMERIC SUBSTRATE MATERIALS

Polycarbonates. Currently, all audio CDs (CD-AD), all CD-ROM, and the biggest fraction of substrate disks for WORM and EOD worldwide are manufactured from a modified bisphenol A-polycarbonate (BPA-PC) (3). In 1991, some 1.3 × 10⁹ compact disks were produced, equivalent to an annual amount of about 35,000 t BPA-PC. WORM and EOD disks are manufactured mainly from BPA-PC for sizes of 5.25 in. and below, and glass for larger form factors (eg, 12 in.), partially also from BPA-PC, and in some cases from aluminum or from cross-linked polymers (epoxy resins) (190).



Modification of BPA-PC for adaptation to the conditions of production of CD and CD-ROM disks, and of substrate disks for WORM and EOD was necessary. BPA-PC standard qualities for extrusion and injection molding have, depending on molecular weight, melt flow indexes (MFI), (according to ISO 1130 ASTM 1238 in g 10 min at 300°C 1.2 kg, between less than 3 g 10 min (viscous types) up to 17 g 10 min. For CDs and optical data storage disks, however, MFI values exceeding 30 g 10 min, and for exceptionally short cycle times (5–7 s) even >60 g 10min are demanded at an injection mass temperature of 300°C (see Table 5).

An increase in MFI can be gained by lowering the molecular weight (MFI line in Fig. 23), but this has to be weighed against a steep decrease in general characteristics, eg, resistance to heat softening and impact resistance (Standard PC line in Fig. 23). Possibilities therefore had to be found to shift the characteristics line to lower molecular weights. This could be achieved by exchanging the usual phenolic end groups by *p*-alkylphenols with branched alkyl groups. For the so-called CD qualities, the MFI could be increased to values around 66 g 10 min by lowering the molecular weight from about 30,000 to about 18,000, this without a significant loss in characteristics (CD-modified PC line in Fig. 23) (191,192).

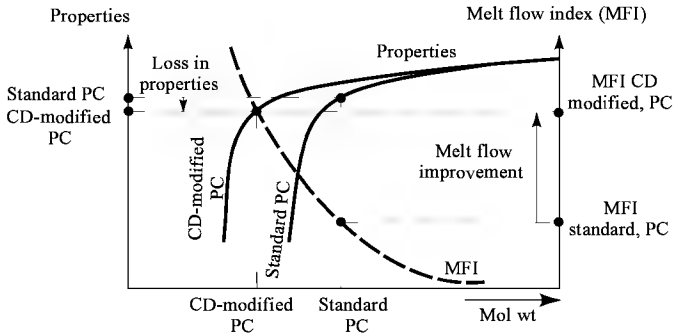


Fig. 23. Correlation between properties (general characteristics), melt flow index (MFI), and mol wt for standard BPA polycarbonate and CD-modified BPA polycarbonate (191,192).

At the same time, a decrease in orientation- and stress-birefringence for the modified BPA-PC could be achieved by optimizing the processing conditions (184,187,189,190,193); this lowers birefringence below 20 nm mm (single path) for injection-printed injection molded CDs and substrate disks (Fig. 24) (191). The typical characteristics of a modified BPA-PC especially developed for CD production are listed in Table 6 (187,193).

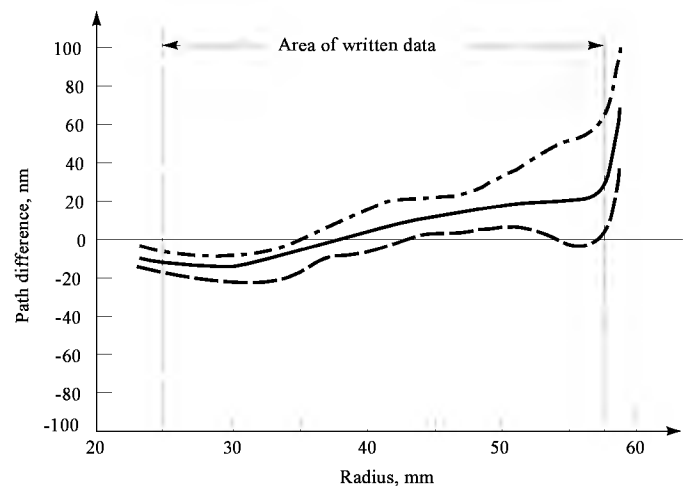


Fig. 24. Birefringence (path difference) of a compact disk, made from CD-modified BPA polycarbonate (191). (○), max value; (◊), min value; (—), mean value.

Table 6. Characteristic Properties of CD-Modified BPA Polycarbonate^a

Properties	Standards	Value
<i>Optical properties</i>		
refractive index	ASTM D542	1.583
light transmission ($\lambda = 800$ nm); 3.2 mm, %	ASTM D1003	92
birefringence (single path), nm/mm		<40 ^b
<i>Thermal/mechanical properties</i>		
vicat softening temperature VST ^c B 120, °C	ASTM D1525	141
heat deflection temperature HDT ^c A, B, °C	ASTM D648	120–134
tensile stress at yield, MPa ^c	ASTM D636	62
elongation at break, %	ASTM D638	>50
impact strength (notched Izod; 3.2 mm), J/m ^d	ASTM D256	410
<i>Physical/chemical properties</i>		
water absorption (immersion), %	ASTM D570	<0.35
water absorption (at 23°C, 50% rh), %	ASTM D570	<0.15
<i>Rheological properties</i>		
melt flow index (300°C/1.2 kg), g/10 min	ASTM D1238	70
<i>Processing properties</i>		
cycle time, s		7 ^e

^a Makrolon CD 2005, Bayer AG (Germany) and Mobay Corp. (United States) (193).

^b With optimal processing; <20 nm/mm.

^c To convert MPa to psi, multiply by 145.

^d To convert J/m to ft-lbf/in., divide by 53.38.

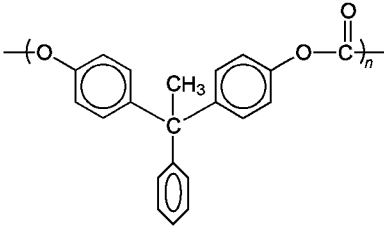
^e With optimal processing; <5 s.

Although CD-modified BPA polycarbonate can be employed without problems for CD-DA and CD-ROM, the use as a substrate material for EOD(MOR) requires an optimum selection and meticulous adherence to production conditions to achieve the required birefringence values of less than 20 nm/mm (184,193).

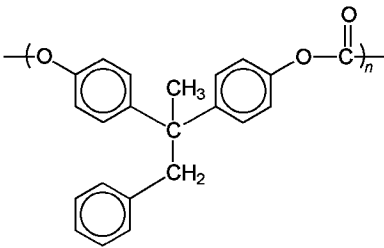
For substrates of WORM and EOD(PCR) disks the industry in the future wants polymers that have a markedly improved resistance to heat softening compared to BPA-PC and, if possible, a lower water absorption and lower birefringence, but otherwise maintain the good characteristics in toughness, production, and cost (194). This goal is being approached in different ways: further modification of BPA-PC, newly developed polymers, improvement of the processing characteristics of uv-curable cross-linked polymers, and development of special copolymers and polymer blends, eg, TMBPA-PC, SAN and PPE-PS.

Polycarbonates with Improved Heat Resistance. The high optical anisotropy of BPA-PC is mainly caused by the strong polarizability of the aromatic rings in the backbone. This inherent high birefringence of PC can be reduced by the installation of groups with high birefringence (eg, phenyl groups) in lateral side groups (195,196).

The birefringence for phenyl-substituted PC (**4**) ($T_g = 176^\circ\text{C}$) is reduced to about 50%, for benzyl substituted PC (**5**) ($T_g = 138^\circ\text{C}$) to about 25%, and for four-ring bisphenol PC (**6**) to 8% of the value for BPA-PC (183,190,195,197,198) on condition of an optimum conformation of the phenyls in the side groups perpendicular to the aromatic rings in the backbone. In reality, however, these low birefringence values are not achieved, because the optimum conformation of the phenyl rings cannot be achieved in injection-stamped disks.

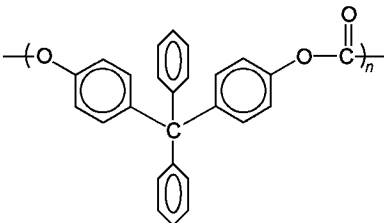


(4)



(5)

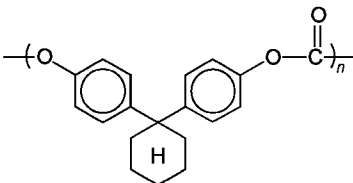
Regarding the glass-transition temperature, only the four-ring bisphenol-PC (6) exhibits a remarkable increase of T_g (220°C) over BPA-PC, but the high brittleness prohibits its use in practical applications (195,196).



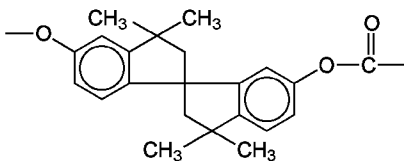
(6)

Numerous attempts have been made to reach improved products by installing phenyl side groups on the aromatic rings of the backbone; for this, sometimes the isopropylidene group has been partially or wholly substituted. The corresponding products exhibit a greatly reduced birefringence, but their resistance to heat softening leaves something to be desired (199). The same holds true for polycarbonates with *t*-butyl and isopropyl substituents on the rings of the backbone (200).

Another path has been followed by completely changing the isopropylidene group. Replacement by a side-based cyclohexyl ring by itself raises the glass-transition temperature markedly (179°C) (7) (201); the additional lowering of birefringence by *t*-butyl or phenyl substituent groups on the aromatic rings of the backbone, however, results in overall reduction of T_g , 152°C for *t*-butyl; 161°C for phenyl (202).



(7)



(8)

A decisive breakthrough to a T_g range above 230°C was achieved with a polycarbonate based on spirobisindan (SBI-PC, (8)) (203) or 3,3,5-trimethylcyclohexane (TMC-PC, (7) wherein the cyclohexane ring bears three methyl groups) (204).

The nearly symmetrical composition of SBI-PC ($T_g = 230^\circ\text{C}$) makes birefringence disappear in homopolymers, but the material becomes very brittle due to the blocking of the free rotation of the aromatic rings, which puts its technical application in question. Only a copolymerization with 80 wt % BPA-PC reaches sufficient levels of impact resistance; but T_g is lowered to 170°C and birefringence increases to 80% of that of BPA-PC (195,196). In contrast, TMC-PC as a homopolymer already has sufficient impact resistance at a T_g of 238°C and a birefringence of 83% of that of BPA-PC (195,205) (Table 7).

Table 7. Properties of TMC-Polycarbonate and its Copolymers with BPA-Polycarbonate^a

Property	BPA-PC:TMC-PC				
	100:0	80:20	65:35	45:55	0:100
glass transition temperature, T_g , °C	149	174	187	205	238
heat resistance HDT-B (0.45 N/mm ²), ^b °C	138	163	174	195	223 ^c
impact strength (ISO 180), ^d kJ/m ^{2e}	not broken	not broken	not broken	not broken	^f
notched impact strength, kJ/m ^{2e}	45	14	8	<8	<5
melt viscosity (at 360°C and shear rate 1000 s ⁻¹), Pa·s ^g	120	170	220	280	380
max birefringence $\Delta n \times 10^3$	2.06	1.99 ^c	1.93 ^c	1.86 ^c	1.70

^a Ref. 206.

^b To convert N/mm² to psi, multiply by 145.

^c Interpolated data.

^d nb not broken.

^e To convert kJ m^{-2} to ftlb f in^{-2} , divide by 2.10.
^f From 10 measurements of 0:10, eight are nb and two $> 150 \text{ kJ m}^{-2}$.
^g To convert Pa s to P , multiply by 10.

Copolymers and Blends of PC. Numerous co- and terpolymers as well as polymer blends of BPA-PC have been developed and their suitability as substrate materials for optical data storage media has been tested (Table 8) (195). From these products, three lines of development have been chosen for closer examination.

Table 8. Substrate Materials for Optical Data Storage

Substrate material	Manufacturer
<i>Copolymers</i>	
copolymer of PC and styrene (24:76)	Idemitsu
graft copolymer of PC and styrene	Mitsubishi
copolymer of PC with <i>o,o'</i> -biphenyl units, $T_g \sim 160^\circ\text{C}$	Toray
copolymer of PC with dihydroxytricyclodecane units, $\text{HDT} \sim 135^\circ\text{C}$	Tosoh
copolymer of PC and spirobisindane-PC (SBI-PC)	General Electric
copolymer of PC and trimethylcyclohexane-PC (TMC-PC)	Bayer AG
<i>Blends</i>	
PC with modified polystyrene (60:40)	Sumitomo
PC with MMA-styrene copolymer ($\text{LCST}^a \sim 240^\circ\text{C}$)	Mitsubishi
PC with SMA copolymer	Nippon Steel
PC with dimethyl BPA-PC	Bayer AG
PC with four-ring bisphenol-PC	Bayer AG
PC with trimethylcyclohexane-PC (TMC-PC)	Bayer AG
PC-cocondensates (BPA-PC, benzyl-substituted PC, four-ring bisphenol-PC, PC with lateral phenyl side groups on the aromatic rings)	Bayer AG
TMBPA-PC with polystyrene (62:38, $\text{LCST}^a \sim 240^\circ\text{C}$)	Philips
TMBPA-PC with modified PS	Tosoh
TMBPA-PC with modified SAN	Bayer AG

^a LCST = lower critical solution temperature.

Copolymers and Blends of BPA-PC and Modified PS. Theoretically, a blend or copolymer of 60 parts BPA-PC (positively birefringent) and 40 parts PS (negatively birefringent) should yield a product free of birefringence (Fig. 25) (207). In spite of modifications to PC to improve the compatibility, no blend could be produced which would be optically isotropic and thus suitable as a substrate material. The same holds true for PC-PS-copolymers (208) in which PC:PS = 77 : 33.

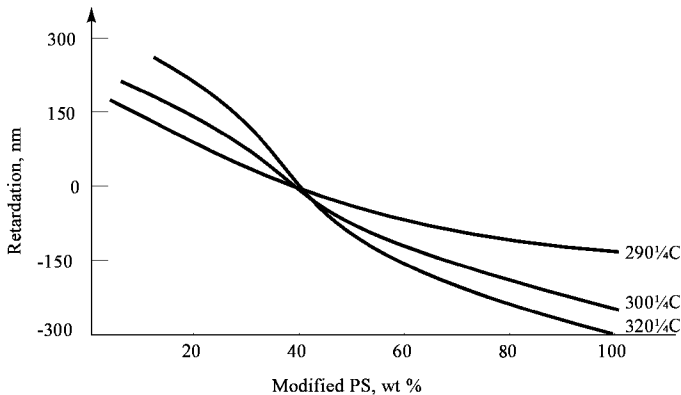
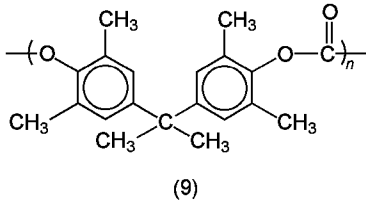


Fig. 25. Influence of blend composition on birefringence (207). Blend is PC and modified PS. Temperatures noted are molding temperatures.

Blends of Tetramethylbisphenol A-PC (TMBPA-PC) with Modified PS or Styrene-Acrylonitrile(SAN) Copolymer. By installing halogen atoms on the aromatic rings of the PC-backbone, not only the resistance to heat softening can be increased (eg, TMBPA-PC: $T_g \sim 203^\circ\text{C}$) (209), but also the compatibility with olefins. Optically homogenous blends have been described (210) of TMBPA-PC (positively birefringent) with modified PS or modified SAN (both negatively birefringent) (see Table 8); due to their low birefringence and increased resistance to heat softening these products have been recommended as substrate materials for magneto-optical data storage (197,198).



Copolymers of BPA-PC with TMC-PC. These copolymers (206) (APEC-HT, Bayer AG, Leverkusen, Germany) have high resistance to heat softening, sufficient impact resistance, lower birefringence than BPA-PC, and low melt viscosity for easy processing. The most important data of selected copolymers was shown in Table 7 (195,206). These values reveal the material's fitness as a basis material for optical data storage disks especially when resistance to heat softening is required which is markedly higher than that of BPA-PC.

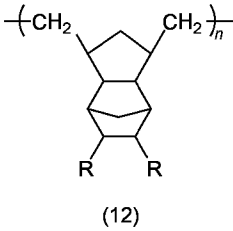
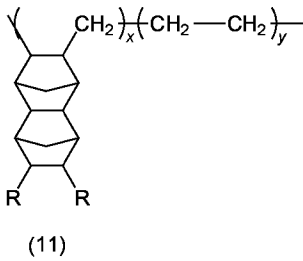
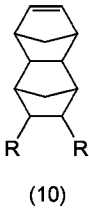
Poly(methyl methacrylate). PMMA offers distinct advantages over BPA-PC with respect to significantly lower birefringence, higher modulus, and lower costs, but has not been successful as a material for audio CDs and CD-ROM as well as a substrate material for WORM and EOD disks because of its high water absorption (which makes it prone to warp) and its unsuitability for metallizing, and less so because of its low resistance to

heat softening. Only 12 in. video disks are manufactured from PMMA by injection molding or the 2P process (photopolymerization process), because the glued sandwich arrangement in this product prevents warp due to high water absorption.

In the literature (182,189,211) can be found numerous studies which tried to alleviate these disadvantages by modification, copolymerization, or blending with proper partners: MMA-CHMA copolymers (methyl methacrylate-cyclohexyl methacrylate) have a lower water absorption than pure PMMA but at the price of a reduction in strength (182,189). MMA-St copolymers (methyl methacrylate-styrene) also show lower water absorption but have a higher birefringence (182). For PMMA-PVC blends (212) and PMMA-PVDF blends (213) (PVC = poly(vinyl chloride), PVDF = poly(vinylidene fluoride)) in the right blend mixture, birefringence is indeed canceled out, but the other properties (eg, resistance to heat softening) do not meet the requirements.

More recently, test products were created of a blend of PMMA with a phenyl-substituted methacrylate; these products have a glass-transition temperature of around 125°C, a significantly reduced water absorption compared to pure PMMA of about 0.32%, but also a higher birefringence (a stress-optic coefficient of 5.2×10^{-11} , compared with 0.3×10^{-11} for PMMA and 6.8×10^{-11} for BPA-PC).

Cyclic Polyolefins (CPO) and Cycloolefin Copolymers (COC). Japanese and European companies are developing amorphous cyclic polyolefins as substrate materials for optical data storage (213–217). The materials are based on dicyclopentadiene and/or tetracyclododecene (10), where R = H, alkyl, or COOCH₃. Products are formed by Ziegler-Natta polymerization with addition of ethylene or propylene (11) or so-called metathesis polymerization and hydrogenation (12), (101,216). These products may still contain about 10% of the dicyclic structure (216).



The production of CPO is based on relatively inexpensive cyclic substances; these must be derivatized, however, to meet the requirements of resistance to heat softening and suitability for metallization. Metathesis polymerization is problem-prone, since relatively large amounts of catalyst (WCl₆, C₂H₅AlCl₂) must be removed by solvent extraction (216). In the process, the price of CPO, at small batches, is several times higher than that of BPA-PC.

The principal advantages of CPO over the current substrate materials based on BPA-PC are very low water absorption (depending on type less than 0.1%), resistance to polar solvents and to acids and bases, low birefringence (comparable to that of PMMA), and short vacuum time (for sputtering ferromagnetic layers). Disadvantages include high proneness to warp, low impact resistance, significantly higher melt viscosity (longer cycle time in disk production), unsatisfactory metallizability, and high expense.

Table 9 compares the most important properties of substrate materials based on BPA-PC, PMMA, and CPO (three different products) (216,217). The future will prove if the current disadvantages of CPO against BPA-PC regarding warp, processibility (melt viscosity), and especially cost can be alleviated. Cyclic polyolefins (CPO) and, especially cycloolefin copolymers (COC) (218) and blends of cycloolefin copolymers with suitable engineering plastics have the potential to be interesting materials for substrate disks for optical data storage.

Table 9. Comparison of Characteristic Properties of Substrate Materials

Properties	BPA-PC ^a	PMMA ^b	CPO ^c	CPO ^c	CPO ^c
light transmission, %	>90	>90	>90	>90	92
birefringence, nm/mm	<40	<20	<25	<20	<20
glass-transition temperature, °C	145	105	140	171	150
tensile stress at yield, MPa ^d	62	73	64	42	59
elongation at break, %	>50	5	10	16	3
impact strength, notched kJ/m ² ^e	20–30	2	4	3	
water absorption, (100% rh), %	<0.35	2.1	<0.01	0.2	0.01
melt flow index (280°C), g/10 min	57		15		
vacuumizing time, min ^f	78	200	14		

^a CD-modified BPA-polycarbonate (CD 2005, Bayer AG).

^b Poly(methyl methacrylate).

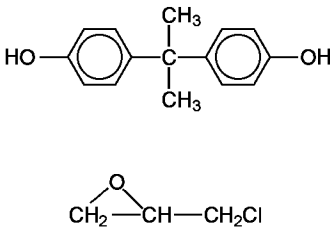
^c Cyclopolyolefins.

^d To convert MPa to psi, multiply by 145.

^e To convert kJ/m² to ft-lbf/in.², divide by 2.10.

^f Mitsui method

Cross-Linked Polymers. In the 1980s, not only glass and BPA-PC but also uv-curable cross-linked polymers, eg, epoxy resins, were used as substrate material for optical mass storage disks with large diameters (12 in., 14 in.) (219). The epoxy resins consisted of compounds containing one or several highly reactive epoxy or hydroxyl groups. The common epoxy resins (EP) mainly are reaction products of bisphenol A and epichlorohydrin:



The resulting bisepoxy compounds are cross-linked cold with polyamines, if necessary with added accelerators. A hot cure can either be accomplished with amines or anhydrides (eg, phthalic acid anhydride). If suitable initiators are present, EP systems can also be cross-linked by radiation. Special, uv-curable epoxy resins (qv) for substrate disks for optical data storage (Sumitomo Bakelite, Toshiba) excel by means of their very low birefringence (<5 nm mm) and high Young's modulus. Resistance to heat softening and water absorption are similar to BPA-PC, but impact resistance is as low as that of PMMA.

The decisive disadvantage for industrial mass production of substrate disks is the extremely long curing time (eg, 5 h at 100°C); thus efforts to reduce curing time have been numerous. It is possible to produce substrate disks with up to 30-cm diameter by using uv-curing reaction resins based on unsaturated carbon acid esters and other vinyl compounds. The cycle times are as low as 15 to 100 s when an additional post-cure, as necessary for all cross-linked polymers with high *T_g*, ranging from 1 to 5 h at 60 to 100°C, is applied (220).

Table 10 compares the values of different experimental uv-curable cross-linked polymers with those of BPA-PC for the most important properties of substrate materials (220). In spite of this remarkable progress in the development of fast curing cross-linked polymers, BPA-PC and, to a small extent, glass are still the materials of choice for substrates for optical data storage.

Table 10. Comparison of Characteristic Properties of CD-Modified BPA-Polycarbonate with UV-Curing Duromer

Properties	BPA-polycarbonate ^a	UV-curing duromers ^b
refractive index, <i>n</i> ^D mn	1583	1.55–1.58
birefringence ^c , nm mm		
M ₁	<40	0–10
M ₂	500–800	60–100
light transmission, %	90–92	90–92
glass-transition temperature, °C	145	120–200
water absorption (100% rh), %	0.35	0.36–0.13
coefficient of linear thermal expansion, 10 ^{−6} K	51–71	60–75
tensile modulus, N mm ^{2d}	2300	3000–4000
impact strength, kJ m ^{2e}	not failure	20–4
release time setting time, s	7	100–15 (uv) ^f

^a CD-modified BPA-polycarbonate (CD 2005, Bayer AG (Germany) and Mobay Corp. (United States)).

^b Duromers (cross-linked polymers) based on highly reactive resins with short setting times.

^c Laboratory conditions: M₁ (direction perpendicular to disk plane) < 20 nm mm; M₂ (parallel to plane) < 500 nm mm.

^d To convert N mm² (MPa) to psi, multiply by 145.

^e To convert kJ m² to ftlbf in.², divide by 2.10.

^f Additional (60–120°C) ca 1–5 h.

Other Polymers. Besides polycarbonates, poly(methyl methacrylate)s, cyclic polyolefins, and uv-curable cross-linked polymers, a host of other polymers have been examined for their suitability as substrate materials for optical data storage, preferably compact disks, in the last years. These polymers have not gained commercial importance: polystyrene (PS), poly(vinyl chloride) (PVC), cellulose acetobutyrate (CAB), bis(diallylpolycarbonate) (BDPC), poly(ethylene terephthalate) (PET), styrene–acrylonitrile copolymers (SAN), poly(vinyl acetate) (PVAC), and for substrates with high resistance to heat softening, polysulfones (PSU) and polyimides (PI).

Reduction or even complete compensation of birefringence by mixing polymers with positive birefringence (PC, PVC, PETP, PPE, PVDF, etc) with polymers with negative birefringence (PMMA, PS, PAN, etc) has been the consistent strategy.

Newer developments involve poly(4-methyl-1-pentene) (TPX), PS or PPE blends, and block copolymers.

The TPX experimental product of Mitsubishi Petrochemical Ind. (221) is an amorphous, transparent polyolefin with very low water absorption (0.01%) and a glass-transition temperature comparable to that of BPA-PC (ca 150°C). Birefringence (<20 nm mm), flexural modulus, and elongation at break are on the same level as PMMA (221). The vacuum time, the time in minutes to reach a pressure of 0.13 mPa (10^{−6} torr), is similarly short like that of cyclic polyolefins. Typical values of TPX are listed in Table 11. A commercial application of TPX is not known as of this writing.

Table 11. Comparison of Different Substrate Materials^a

Property	Test method	CD-modified BPA-PC	TMC	BPA-PC ^b
density, g cm ³	ASTM D792	1.2	1.2	
light transmission, %	ASTM D1003	> 90		90
refractive index	ASTM D542	1.583	1.565	
Abbe number, $\frac{n-1}{n_{bl}-n_{red}}$		30		~ 30
birefringence, nm mm	single path	< 40		< 35
glass-transition temperature, °C	DSC method	145		205
heat defl temperature, HDT A, °C	ASTM D648	129		179 197 ^f

bending modulus, MPa ^g	ASTM D720	2.400	2.200
elongation, at break, %	ASTM D636	> 50	50
flatness (tilt), mrad		< 5	< 5
impact strength (Izod), kJ m ⁻²	ISO 180 1C	not broken ^h	not broken ^h
notched impact strength (Izod), kJ m ⁻²	ISO 180 1A	20–30	5–8
melt flow index, g 10 min (280°C)	ASTM D1238	57	~ 30
vacuumizing time, min (1.3 mPa ^l)	Mitsui method	78	78
water absorption, % 23°C, 50% rh	ASTM D570	0.15	0.15
water permeation, g m ² 24 h		1.53	1.53
processing ease		+	+

PMMA	uv-DM ^c	CPO ^d	CPO ^d	PS PPE (71:29)	TPX ^e	Glass
1.19		1.01	1.08		1.05	2.5
>90	> 90	>90	>90	>92	>92	>92
1.49		1.53	1.51	1.59	1.55	1.5
58		52	57			50
<20	< 10	<25	<20	<20	<20	≪5
105	120...200	140	171	134	150	
90		123			140	
3.000	~ 3.000	2.400			3.000	
5		10	16		~5	
<5		<5	<5			<2
18	4–20	22	36			
2	<2	4	3			
		15				
200		14				
1.2	0.36...0.13	<0.01	0.2	0.1...0.4	0.01	0
1.14		<0.01				0

^c Ultraviolet-curing cross-linked polymer.

^d Cyclopolyolefin (two different products).

^e Poly(4-methyl-1-pentene).

^a Data from manufacturer's literature.

^b TMC-PC:BPA-PC = 55:45 (Apec-HT, Bayer AG (Germany)).

^f Annealed for 2 h.

^g To convert MPa to psi, multiply by 145.

^h nb not broken.

¹ To convert kJ m⁻² m² to fdlbf in.², divide by 2.10.

^l To convert mPa to torr, divide by 1.3 × 10².

Experimental products with a mass ratio of polystyrene to poly(phenylene ether) (PS:PPE = 65 : 35) exhibit glass-transition temperatures of 136°C and water absorption of 0.1% (according to other measurements, 0.4%) together with excellent transparency (>90% light transmission); birefringence in cast test samples (without orientation of molecules due to injection molding) is completely compensated (222,223). Typical values of this blend are also listed in Table 11. Interestingly, blend composition for no birefringence depends strongly on the molecular weight of the individual components (223). In injection molded test samples, a strong dependence of birefringence on the processing conditions could be seen. This product also is still waiting for a commercial application. Calculations of Sumitomo Chemicals show that only an annual production upward of 1000 to 2000 t (corresponding to about 30 to 60 × 10⁶ substrate disks of 5.25 in. dia) would be economical (224).

It has been reported that block copolymers with appropriately chosen partners and mixing ratios yield materials suitable for use in substrate disks for optical data storage. An example is polyarylate–polystyrene block copolymer with a PS content between 40 and 60% (225).

Comparison of Substrate Materials

For CD-DA (CD-digital audio) and CD-ROM including all variants, a CD-modified BPA-PC is used exclusively. BPA-PC is stipulated in the specifications for CD-DA.

WORM disks with diameters of 130 mm (5.25 in.) and 200 mm (8 in.) are manufactured almost exclusively from modified BPA-PC, including Kodak's Photo-CD. Glass is used predominantly only for WORM disks with 300 mm (12 in.) diameter.

For EOD disks with 86 mm (3.5 in.) diameter and for Sony's Mini-Disk (67 mm diameter), modified BPA-PC is used exclusively; for EOD disks with 130 mm diameter, modified BPA-PC (mainly) as well as glass (eg, Philips, Hitachi) are used.

In spite of the decision of commercial suppliers of digital optical storage disks, it is of interest to compare the most important properties of glass and polymers and rate them according to their characteristics relevant to optical storage disks. For a direct visual comparison, Figure 26 depicts the most important properties relevant to data storage in spider charts. The innermost circle symbolizes the most unfavorable case of each property and the outermost circle the most favorable case (187). Table 11 presents the numeric values available thus far; if varying numbers exist, a qualitative assessment was made (according to Ref. 216).

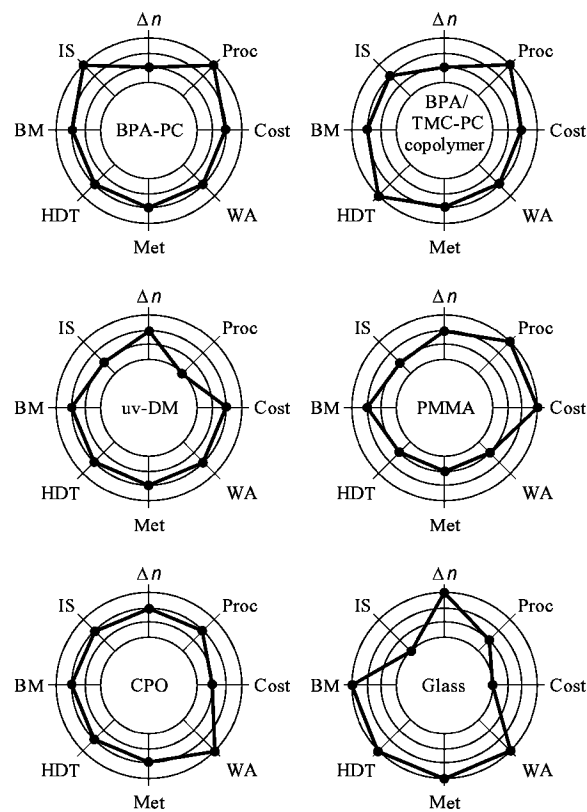


Fig. 26. Qualitative comparison of substrate materials for optical disks (187): Δn – birefringence; IS – impact strength; BM – bending modulus; HDT – heat distortion temperature; Met – metallizability; WA – water absorption; Proc – processibility. The materials are bisphenol A–polycarbonate (BPA-PC), copolymer (20:80) of BPA-PC and trimethylcyclohexane–polycarbonate (TMC-PC), poly(methyl methacrylate) (PMMA), uv-curable cross-linked polymer (uv-DM), cyclic polyolefins (CPO), and, for comparison, glass.

Of practical interest are detailed studies to influence the magneto-optical properties of RE-TM materials by the substrate material and the substrate adhesion of RE-TM layers by the selected deposition technique (226). Accordingly, measurements have been performed on glass, BPA-polycarbonate, and poly(ethylene terephthalate) (as a flexible substrate).

Future Developments

The question as to whether and to what extent and in what area optical mass storage would replace magnetic systems (disk, tape) was controversially being discussed in the 1980s. In spite of all predictions of an imminent substitution, as of late 1994 magnetic hard disks still are the system of choice for computer-dedicated mass storage due to their speed (access time, transfer rate), physical size, and energy consumption; this is especially true when memory-intensive applications are run which use the hard disk as virtual memory.

There is no competitive situation for data storage disks with embossed information (CD-ROM) and recordable nonerasable disks (WORM); no counterpart to CD-ROM and WORM exists among magnetic memories. EOD drives are best compared to floppies and removable hard disk media given their possibility of easy and problem-free disk exchange and a capacity on the order of that of removable magnetic media (Tape, Bernoulli, SyQuest).

Generally, magnetic media and optical media do not compete; they complement each other due to their specific advantages. It can be expected that also in the future both system families will not stand in opposition to each other, but will and have to solve the upcoming problems of data storage together according to their specific characters and advantages.

The acceptance of optical data storage into the mass storage market, which is as yet exclusively dominated by magnetic systems, will be fundamentally boosted if optical drives and media are subject to uniform standards and become fully compatible, and multiuser drives are offered which enable the user to employ alternatively CD-ROM and EOD disks, and maybe WORM disks as well (and CD-R disks, respectively). A prerequisite, however, will be whether rewritable optical memories will use the MOR or the PCR process. This accord especially will be hard to reach.

Generally it can be said that optical systems will assume an ever increasing market share (depending on the achievement of uniform standards) of the data storage market which is currently dominated by magnetic systems. Additionally they will advance into new applications. Up to the end of the twentieth century, complementary technologies rather than a conflict between optical and magnetic mass memories are likely.

BIBLIOGRAPHY

1. G. Kaempf, H. Loewer, and M. W. Witman, *Polym. Eng. Sci.* **27**, 1421 (1987); G. Kaempf, *Ber. Bunsenges. Phys. Chem.* **89**, 1179 (1985).
2. M. Hartmann, J. Braat, and B. Jacobs, *IEEE Trans. Magn.* **MAG-20**, 2013 (1984).
3. M. Emmelius, G. Pawlowski, and H. W. Vollmann, *Angew. Chem.* **101**, 1475 (1989).
4. K. Y. Law, P. S. Vincett, and G. E. Johnson, *Appl. Phys. Lett.* **39**, 718 (1981); D. G. Howe and J. J. Wrobel, *J. Vac. Sci. Technol.* **18**, 92 (1981).
5. V. Novotny and L. Alexandru, *J. Appl. Phys.* **50**, 1215 (1979), and *J. Appl. Polymer Sci.* **24**, 1321 (1979).
6. M. C. Gupta, *Appl. Opt.* **23**, 3950 (1984).
7. P. Kivits, R. de Bont, and P. Zalm, *Appl. Phys.* **24**, 273 (1981); P. Kivits and co-workers, *Thin Solid Films* **87**, 215 (1982); J. E. Kuder, *J. Imaging Technol.* **12**, 140 (1986).
8. M. Matsuoka, *Mol. Cryst. Liq. Cryst. Sci. Technol., Sect. A* **224**, 85 (1993); A. Kakuta, in M. Matsuoka, ed., *Infrared Absorbing Dyes*, Plenum Press, New York, 1990, H. T. Macholdt, *Chem. Umserev teit* **24**(4), 176 (1990); P. F. Foudon, in D. R. Waring, ed., *Chemical Applied Dyes*, Plenum Press, New York, 1990, p. 381; F. Jones, *Rev. Prog. Color. Relat. Top.* **19**, 20 (1989).
9. V. P. Jipson and C. R. Jones, *J. Vac. Sci. Technol.* **18**, 105 (1981), and *IBM Techn. Discl. Bull.* **24**(1A), 298 (1981).
10. G. Kaempf, H. Loewer, and M. Witman, *Kunstst./German Plastics* **76**, 1077 (1986); M. Witman, L. Loewer, and G. Kaempf, *SPI/SPE Plastics-West*, 40 (1987).
11. D. J. Gravesteijn and J. van der Veen, *Philips Techn. Rev.* **41**, 325 (1983–1984); D. J. Gravesteijn, C. Steenbergen, and J. van der Veen, *Proc. SPIE Int. Soc. Opt. Eng.* **420**, 327 (1983).
12. R. S. Jones and J. E. Kuder, in K. L. Mittal, ed., *Proc. Amer. Chem. Soc. Symp. on Polymers in Information Storage Technology*, Los Angeles, Calif., Sept.

25–30, 1988, Plenum Press, New York, 1989, p. 3.

13. **Eur. Pat. A 254,553** (1988), S. Hayashida and S. Tai (to Hitachi Chem. Co., Ltd.), **Eur. Pat. A 279,426** (1986), R. S. Jones, D. E. Nikles, and M. E. Kenney (to Celanese Corp.), **Eur. Pat. A 296,876** (1988), T. Seiji and co-workers, (to Hitachi Chem. Co., Ltd.).

14. T. Kanno, H. Watanabe, and S. Nozaki, *Proc. 34th Conf. Soc. Appl. Phys.*, 720 (1987); **Jpn. Pat. 62–223, 287** (1987), K. Morikawa, H. Shibano, and H. Yamazaki, (to Kao Corp.); **Jpn. Pat. 63–57, 287** (1988), S. Hirose and co-workers, (to Mitsui Toatsu Chem. Inc.).

15. **Ger. Pat. A 3,711,762** (1988) B. Albert, H. Kuppelmaier, and G. Wagenblast (to BASF AG).

16. M. Ito and co-workers, *Proc. SPIE Int. Soc. Opt. Eng.* **420**, 332 (1983); K. Takagi and co-workers, *Dyes Pigm.* **6**, 177 (1985); S. H. Kim and co-workers, *Chem. Express* **2**, 73 (1987); S. H. Kim, M. Matsuoka, and T. Kitao, *Chem. Lett.*, 1351 (1985).

17. H. W. Vollmann, *J. Inf. Rec. Mater.* **20**, 3 (1992).

18. P. Kullik, ed., *Optische Speicherplattensysteme*, Teil 1, Arbeitsgemeinschaft für wirtschaftliche Verwaltung e.V., Schrift 448, 1988.

19. B. G. Huth, *IBM J. Res. Develop.* **18**, 100 (1974).

20. P. Hansen, *J. Appl. Phys.* **63**, 2364 (1988).

21. M. Mansuripur, *J. Appl. Phys.* **66**, 6175 (1989).

22. S. Klahn and D. Raasch, *IEEE Trans. Magn.* **MAG-26**, 1918 (1990); R. A. Hajjar and H-P. D. Shieh, *J. Appl. Phys.* **68**, 4199 (1990).

23. S. Tsunashima, H. Tsuji, and S. Uchiyama, *J. Magn. Soc. Jpn.* **5**, 973 (1981).

24. C. J. Lin, *J. Appl. Phys.* **67**, 4409 (1990).

25. S. Ohnuki and co-workers, *Moris 91*, 18-K-03 (1991); *J. Magn. Soc. Jpn.* **15**, (S1), 399 (1991).

26. N. Ohta and co-workers, *SPIE* **1663**, 232 (1992).

27. H. Kronmüller, *Phys. Stat. Sol.* **144(b)**, 385 (1987).

28. R. J. Gambino, P. Chaudhaari, and J. J. Cuomo, *AIP Conf. Proc.* **18**, 578 (1974).

29. J. Orehotsky and K. Schröder, *J. Appl. Phys.* **43**, 2413 (1972).

30. P. Chaudhaari, J. J. Cuomo, and R. J. Gambino, *Appl. Phys. Lett.* **22**, 337 (1973).

31. P. Hansen, in K. H. J. Buschow, ed., *Handbook of Magnetic Materials*, Vol. 6, Elsevier Science Publishers B.V., New York, 1991, Chapt. 4.

32. X. Hu and co-workers, *Proc. Intermag '93*, Stockholm, 1993.

33. P. Hansen and H. Heitmann, *IEEE Trans. Magn.* **25**, 4390 (1989).

34. C. J. Lin, *Mat. Res. Soc. Symp. Proc.* **150**, 15 (1989).

35. F. J. A. M. Greidanus and W. B. Zeper, *MRS Bulletin* **15**(4), 31 (1990).

36. D. Mergel, P. Hansen, and S. Klahn, *Mat. Sci. Forum* **62–64**, pt II, 547 (1990).

37. M. Hartmann and co-workers, *J. Mag. Soc. Jpn.* **15**(S1), 165 (1991).

38. D. Mergel and co-workers, *NATO ASI Series B* **259**, 249 (1991).

39. D. Mergel, P. Hansen, and D. Raasch, *Proc. SPIE 1663, Optical Data Storage 1992*, 240 (1992).

40. D. J. Sellmyer, R. D. Kirby, and S. S. Jaswal, "The Magnetism of Amorphous Metals and Alloys", in J. A. Fernandez-Baca and W. Y. Ching, eds., World Scientific Publishing, Ltd., 1993.

41. P. Hansen and co-workers, *J. Appl. Phys.* **66**, 756 (1989).

42. P. Hansen and co-workers, *J. Appl. Phys.* **69**, 3194 (1991).

43. D. H. Ryan and co-workers, *Phys. Rev. B* **40**, 11208 (1989).

44. T. Suzuki, C. J. Lin, and A. E. Bell, *IEEE Trans. Magn.* **MAG-24**, 2452 (1988).

45. D. Mergel, H. Heitmann, and P. Hansen, *Phys. Rev. B* **47**, 882 (1993).

46. V. G. Harris and co-workers, *Phys. Rev. Lett.* **69**, 1939 (1992).

47. F. Hellman and E. M. Gyorgy, *Phys. Rev. Lett.* **68**, 1391 (1992).

48. R. Sato, N. Saito, and Y. Togami, *Jpn. J. Appl. Phys.* **24**, 1266 (1985).

49. Y. Aoki and co-workers, *IEEE Trans. Magn.* **MAG-21**, 1624 (1985).

50. C. J. Lin and A. E. Bell, *Fourth Joint MMM/Intermag Conf.*, Vancouver, AE-02, 1988.

51. D. Raasch, *J. Magn. Soc. Jpn.* **17** (S1), 192 (1993).

52. P. F. Garcia and D. Meinhalt, *Appl. Phys. Lett.* **47**, 178 (1985).

53. S. Hashimoto, Y. Ochiai, and K. Aso, *J. Appl. Phys.* **66**, 4909 (1989).

54. S. Hashimoto and Y. Ochiai, *J. Mag. Magn. Mat.* **88**, 211 (1990).

55. F. J. A. M. Greidanus and co-workers, *Appl. Phys. Lett.* **54**, 2481 (1989).

56. J. Ferre and co-workers, *Appl. Phys. Lett.* **56**, 1588 (1990).

57. W. P. Zeper and co-workers, *J. Appl. Phys.* **65**, 4971 (1989).

58. W. P. Zeper, F. J. A. M. Greidanus, and P. F. Carcia, *IEEE Trans. Magn.* **MAG-25**, 3764 (1989).

59. D. Weller and W. Reim, *Mat. Res. Soc. Symp. Proc.* **232**, 71 (1991).

60. P. F. Carcia, S. I. Shah, and W. B. Zeper, *Appl. Phys. Lett.* **56**, 2345 (1990).

61. P. F. Carcia and co-workers, *Appl. Phys. Lett.* **58**, 191 (1991).

62. H. W. van Kesteren and W. B. Zeper, *JMMM* **120**, 271 (1992).

63. D. Raasch, *IEEE Trans. Magn.* **29**, 34 (1993).

64. S. Sumi and co-workers, *J. Magn. Soc. Jpn.* **15**(S1), 365 (1991).

65. S. Hashimoto and co-workers, *JMMM* **121**, 471 (1993).

66. W. P. Zeper and co-workers, *IEEE Trans. Magn.* **28**, 2503 (1992).

67. S. Hashimoto, A. Maesaka, and Y. Ochiai, *J. Appl. Phys.* **70**, 5133 (1991).

68. D. Mergel and co-workers, *SPIE* **1274**, 270 (1990).

69. P. Hansen, J. P. Krumme, and D. Mergel, *J. Magn. Soc. Jpn.* **15**(S1), 219 (1991).

70. J. W. D. Martens and A. B. Voermans, *IEEE Trans. Magn.* **MAG-20**, 1007 (1984).

71. M. Gomi, K. Utsugi, and M. Abe, *IEEE Trans. Magn.* **MAG-22**, 1233 (1986); J. P. Krumme and co-workers, *J. Appl. Phys.* **66**, 4393 (1989).

72. A. Hoh, *Jpn. J. Appl. Phys.* **28**(53), 15 (1989).

73. Y. Kaneko and Y. Sawada, *Jpn. J. Appl. Phys.* **26**(S), 23 (1987).

74. **Ger. Pat. 3,716,392**, Y. Kaneko and F. Ohmi.

75. K. Shono and co-workers, *MRS Sympos. F 3-2*, Materials Research Society, Pittsburgh, Pa., 1989.

76. S. Tsunashima and co-workers, *IEEE Trans. Mag.* **MAG-17**, 2840 (1981).

77. R. J. Gambino, T. S. Plaskett, and R. R. Ruf, *IEEE Trans. Magn.* **24**, 2557 (1988).

78. H. Iiyori and S. Takayama, *J. Appl. Phys.* **69**, 4761 (1991).

79. J. Saito and co-workers, *Jpn. J. Appl. Phys.* **26**(S4), 155 (1987).

80. A. Okamuro and co-workers, *J. Magn. Soc. Jpn.* **15**(S1), 447 (1991).

81. K. Aratani and co-workers, *SPIE* **1499**(ODS) 209 (1991).

82. A. Fukumoto and co-workers, *SPIE* **1499** (Optical Data Storage), 216 (1991).

83. H. Iida and co-workers, *ISOM'89* **28**, B-20 (1989); *JLAP*, **28**, 367 (1989).

84. T. Tokunaga, *J. Appl. Phys.* **67**, 4417 (1990).

85. D. Raasch, *J. Mag. Magn. Mat.* **101**, 202 (1991).

86. K. Aratani and co-workers, *Proc. SPIE* **1078**(ODS), 258 (1989).

87. Y. Muto and co-workers, *J. Magn. Soc. Jpn.* **15**(S1), 311 (1991).

88. S. Tanaka and co-workers, *J. Magn. Soc. Jpn.* **15**(S1), 331 (1991).

89. K. Hayashi and O. Okada, *J. Magn. Soc. Jpn.* **15**(S1), 335 (1991).

90. M. Miyamoto and co-workers, *J. Magn. Soc. Jpn.* **15**(S1), 339 (1991).

91. T. K. Hatwar and D. G. Stinson, *J. Appl. Phys.* **69**, 6439 (1991).

92. A. Nakaoki and co-workers, *J. Magn. Soc. Jpn.* **17**(S1), 363 (1993).

93. T. Tokimaga and co-workers, *J. Magn. Soc. Jpn.* **17**(S1), 357 (1993).

94. T. Fukami and co-workers, *J. Appl. Phys.* **67**, 4415 (1990).

95. K. Tsutsumi, *J. Mag. Magn. Mat.* **120**, 247 (1993).

96. D. Mengel, *J. Appl. Phys.* **69**, 4520 (1991).

97. D. Mengel, *J. Appl. Phys.* **70**, 6433 (1991).

98. T. Kobayashi and co-workers, *JJAP.* **20**, 2089 (1981).

99. D. Mergel, *J. Appl. Phys.* **74**, 4072 (1993).

100. T. Fukami and co-workers, *J. Magn. Soc. Jpn.* **15**(S1), 293 (1991).

101. H. W. Vollmann, *Chem. Ind.*, 68 (Oct. 1989).

102. J. W. M. Biesterbos and co-workers, *Thin Solid Films* **58**, 259 (1979); R. Hasegawa and R. C. Taylor, *J. Appl. Phys.* **46**, 3606 (1975); R. B. van Dover, *J. Appl. Phys.* **59**, 1291 (1986).

103. M. Libera and M. Chen, *MRS Bulletin* **15**(4), 40 (1990).

104. E. M. Engler, *Adv. Mater.* **2**, 166 (1990).

105. A. E. Bell, in M. J. Weber, ed., *Handbook of Laser Science and Technology*, Vol. 5, Part 3, CRC Press, Boca Raton, Fla., 1987, p. 65.

106. M. Takenaga and co-workers, *Proc. Photo-Opt. Inst. Eng.* **420**, 173 (1983).

107. A. W. Smith, *Appl. Opt.* **13**, 795 (1974).

108. B. Ryan, *BYTE*, 289 (Nov. 1990).

109. K. A. Rubin, *J. Magn. Soc. Jpn.* **16**(S1), 127 (1991).

110. K. Nishimura and co-workers, *Jpn. J. Appl. Phys.* **28**(3), 135 (1989).

111. K. A. Rubin, *Mat. Res. Soc. Symp. Proc.* **230**, 239 (1992).

112. M. Chen and A. Rubin, *Proc. Soc. Photo Opt. Instrum. Eng.* **1078**, 150 (1989); M. Chen and co-workers, *Appl. Phys. Lett.* **46**, 734 (1985).

113. H. Yamazaki and co-workers, *Adv. Mater.* **5**, 214 (1993).

114. D. P. Birnie, in D. R. Uhlmann and D. R. Ulrich, eds., *Ultrastructure Processing of Advanced Materials*, John Wiley & Sons, Inc., New York, 1992, p. 475.

115. S. R. Ovshinsky, *J. Non-Cryst. Solids* **141**, 200 (1992).

116. T. Ohta and co-workers, *Jpn. J. Appl. Phys.* **28**(3), 123 (1989); H. Ishibashi, M. Moriya, and T. Ohta, *SPIE* **1663** (Optical Data Storage), 427 (1992).

117. I. Morimoto and co-workers, *SPIE* **1663** (Optical Data Storage), 294 (1992).

118. D. E. Nikles and co-workers, *Proc. Soc. Photo Opt. Instrum. Eng.* **1078**, 43 (1989); P. K. Chan and T. R. Hart, *Appl. Opt.* **28**, 1685 (1989); S. Y. Suh and co-workers, *Proc. Soc. Photo Opt. Instrum. Eng.* **1078**, 189 (1989); H. A. Goldberg and co-workers, *Proc. Soc. Photo Instrum. Eng.* **1078**, 170 (1989).

119. M. C. Gupta, *J. Appl. Phys.* **60**, 2932 (1986).

120. N. E. Iwamoto and J. M. Halter, in Ref. 11, p. 117; W. E. Skiens and G. A. Russell in Ref. 12, p. 133; J. M. Halter and N. Iwamoto, *Proc. Soc. Photo Instrum. Eng.* **899**, 201 (1988); M. A. Lind and J. S. Hartman, *Proc. Soc. Photo Instrum. Eng.* **899**, 211 (1988); J. S. Hartman and co-workers, *Proc. Soc. Photo Instrum. Eng.* **1078**, 80 (1989).

121. G. Kaempf, *Polymer J.* **19**, 257 (1987).

122. *High Tech. Materials Alert*, 2 (Jan. 1985).

123. J. E. Kuder, *J. Imaging Technol.* **12**, 140 (1986); F. E. Luborsky, *Mater. Res. Soc. Symp. Proc.* **80**, 375 (1987); R. P. Freeze, *IEEE Spectrum* **25**, 41 (1988); D. A. Carlin, Y. Tsunoda, and A. A. Jamberdino, *Proc. SPIE Int. Soc. Opt. Eng.*, 899 (1988); G. R. Knight and C. N. Kurtz, eds., *Proc. SPIE Int. Soc. Opt. Eng.*, **1078** (1989).

124. H. Дьтт, *Angew. Chem.* **101**, 427 (1989).

125. H. Menzel, *Nachr. Chem. Tech. Lab.* **39**, 636 (1991).

126. R. Bertelson, in G. H. Brown, ed., *Techniques of Chemistry*, Vol. 3, Wiley-Interscience, New York, 1971; N. S. Allen and J. F. McKellar, eds., *Photochemistry of Dyed and Pigmented Polymers*, Applied Science Publishers, London, 1989.

127. M. Kryszewski and P. Uznanski, *Proc. Int. Symp. PME'89, Tokyo 1989*, Kodansha Publ. Tokyo, 1990, and VCH, Weinheim, 1990, p. 549; M. Gehrtz, C. Bräuchle, and J. Voitlander, *J. Amer. Chem. Soc.* **104**, 2094 (1982); C. Lenoble and R. S. Becker, *J. Phys. Chem.* **90**, 62 (1986).

128. Z. G. Gadund, *J. Polym. Sci. Polym. Lett. Ed.* **6**, 57 (1968).

129. M. Kryszewski, D. Lapiensis, and B. Nadolski, *J. Polym. Sci. Polym. Chem. Ed.* **11**, 2423 (1973); M. Kryszewski and co-workers, *J. Chem. Soc. Faraday Trans. II* **76**, 351 (1980).

130. **Brit. Pat. 2,441,759** (Mar. 13, 1975), H. G. Heller; H. G. Heller, *Chem. Ind. London*, (6), 193 (1978); H. G. Heller, *IEEE Proc.* **130**, 209 (1983); A. P. Glaze and co-workers, *J. Chem. Soc. Perkin Trans.*, 957 (1985).

131. R. J. Hart and H. G. Heller, *J. Chem. Soc. Perkin Trans.*, 1321 (1972); G. A. Delzenne, *Adv. Photochem.* **11**, 2 (1979); H. G. Heller and co-workers, *J. Chem. Soc. Perkin Trans.*, 1487 (1974), *Ibid.*, 571 (1978), *Ibid.*, 202 (1981), *Ibid.*, 315 (1986); *Ibid.*, 1599 (1986); F. Eplattenier, *Swiss Chem.* **7**, 19 (1985).

132. I. Cabrera, V. Krongauz, and H. Ringsdorf, *Angew. Chem.* **99**, 1204 (1987), and *Angew. Chem. Int. Ed. Engl.* **26**, 1178 (1987); Y. Suzuki and co-workers, *Polym Bull.* **17**, 285 (1987).

133. J. F. Cameron and J. M. J. Frechet, *Macromolecules* **24**, 1088 (1991).

134. H.-W. Schmidt, *Angew. Chem. Adv. Mater.* **101**, 964 (1989).

135. M. Eich and H. J. Wendorff, *Makromol. Chem. Rapid Commun.* **8**, 467 (1987); M. Eich and co-workers, *Proc. SPIE Int. Soc. Opt. Eng.* **682**, 93 (1987).

136. I. Cabrera, A. Ditttrich, and H. Ringsdorf, *Angew. Chem.* **103**, 106 (1991).

137. V. P. Shibaev and co-workers, *Polym. Comm.* **24**, 364 (1983); V. P. Shibaev and N. A. Plate, in *Advances in Polymer Science*, 60/61, 173, Springer-Verlag, Berlin, 1984.

138. H. J. Coles and R. Simon, *Polymer* **26**, 1801 (1985); H. J. Coles, *Faraday Discuss. Chem. Soc.* **79**, 201 (1985); M. Eich and co-workers, *Makromol. Chem. Rapid Commun.* **8**, 59 (1987).

139. T. Ikeda and co-workers, *Chem. Lett.*, 1679 (1988); T. Ikeda and co-workers, *Mol. Cryst. Liq. Cryst.* **182 B**, 357 and 373 (1990); T. Ikeda and co-workers, *Ibid.* **188**, 207 and 223 (1990); T. Ikeda and co-workers, *Mol. Cryst. Liq. Cryst.* **188**, 235 (1990).

140. T. Ikeda and co-workers, *Macromolecules* **23**, 42 (1990).

141. M. Eich and J. H. Wendorff, *Mol. Cryst. Liq. Cryst.* **169**, 133 (1989); R. Birenheide and H. J. Wendorff, *SPIE Proc.*, 1213 (1990); K. Andele, R. Birenheide, and H. J. Wendorff, *BMFT-Workshop (Ue???336???ue Werkstoffe)*, Dresden, 1991, *Congress Book*, p. 507, and *Makromol. Chem. Macromol. Sympos.*, (1991), *Congress Book*, p. 44; K. Andele and co-workers, *Makromol. Chem. Rapid Comm.* **10**, 477 (1989).

142. J. Stumpe and co-workers, *J. Inf. Rec. Mat.* **27**, 449 (1994); J. Stumpe and co-workers, *Makromol. Chem. Rapid Commun.* **12**, 81 (1991); Th. Fischev and co-workers, *J. Photochem. Photobiol. A: Chem.* **80**, 453 (1994); L. Laeskev and co-workers, *Mol. Cryst. Liq. Cryst.* **246**, 347 (1994).

143. J. Pinsl, C. Bräuchle, and F. H. Kreuzer, *J. Mol. Electr.* **3**, 9 (1987).

144. S. Hvilsted, F. Andruzzi, and P. S. Ramanujam, *Opt. Lett.* **17**, 1234 (1992).

145. K. Anderle and H. J. Wendorff, *Mol. Cryst. Liq. Cryst.*, 191 (1993), in press.
146. G. Wagenblast, K. Beck, and K.-H. Etzbach, *205th Nat. Meeting ACS*, Denver, Colo., 1993, preprints p. 699, and *Nat. Meeting ACS*, Chicago, Ill. (1993), preprints **34**, 709 (1993); **Ger. Pat. 3,704,146** (1988), G. Wagenblast, K.-H. Etzbach, and B. Hisgen (to BASF AG).
147. K. Ichimura and co-workers, *Makromol. Chem. Rapid Commun.* **10**, 5 (1989); S. Tazuke, *Kokagaku* **12**, 85 (1988).
148. W. Hickel and co-workers, *Makromol. Chem. Rapid Commun.* **10**, 353 (1989); M. Sawodny and co-workers, *Makromol. Chem., Macromol. Symp.* **46**, 217 (1991).
149. K. Ichimura and co-workers, *Langmuir* **4**, 1214 (1988).
150. T. Seki and co-workers, *Macromolecules* **22**, 2305 (1989).
151. U. Wiesner and co-workers, *Makromol. Chem.* **191**, 2133 (1990).
152. T. Seki and co-workers, *2nd Int. Sympos. on Chemistry of Functional Dyes*, Osaka, Japan, 1992, Paper P-042, *Congress Book*, p. 91.
153. U. Wiesner and co-workers, *Liq. Cryst.* **11**, 251 (1992).
154. K. Ichimura and M. Sakuragi, *J. Polym. Sci. Lett. Ed.* **26**, 185 (1988); A. Tshuchida, M. Nakano, and M. Yamamoto, *Proc. Int. Symp. PME'89*, Tokyo 1989, Kodansha Publ., Tokyo, 1990 and VCH, Weinheim, 1990, p. 541; C. Bräuchle and co-workers, *Optics Lett.* **6**, 159 (1981); *Ibid.* **7**, 177 (1982).
155. M. Sisido and co-workers, *Macromolecules* **24**, 3993 (1991).
156. N. Hampp, C. Bräuchle, and D. Oesterhelt, *J. Biophys.* **58**, 83 (1986); N. Hampp and C. Bräuchle, in H. Dürr and H. Bouas-Laurent, eds., *Photochromism*, Elsevier, Amsterdam, 1990, p. 954.
157. R. R. Birge, *Annu. Rev. Phys. Chem.* **41**, 683 (1990).
158. D. Oesterhelt, C. Bräuchle, and N. Hampp, *Quarterly Rev. of Biophys.* **24**(4), 425 (1991).
159. R. Margalit and J. Yu, *IEEE EMBS* **12**, 1717 (1990); C. Bräuchle, N. Hampp, and D. Oesterhelt, *Adv. Mater.* **3**(9), 420 (1991); C. Bräuchle and N. Hampp, *BPS'91*, Bayreuth, Germany, 1991, *Congress Book*, p. 102.
160. N. Hampp, C. Bräuchle, and D. Oesterhelt, *4th Int. Sympos. on Bioelectronic and Electronic Devices*, Miyazaki, Japan, 1992, *Congress Book*, p. 72; N. Hampp, D. Oesterhelt, and C. Bräuchle, *Proc. Holographics'90*, Nürnberg, Germany, *Congress Book*, p. 65; N. Hampp, C. Bräuchle, and D. Oesterhelt, *MJR-Bulletin*, p. 56, Nov. 1992.
161. W. Lahti and D. McCarron, *BYTE* **15**, 311 (1990).
162. General Presentation, Laser Card (Optical Memory Card), Drexler Technology Corp., 1989.
163. G. Kaempf, D. Freitag, and W. Witt, *Angew. Makromol. Chem.* **183**, 243 (1990).
164. L. d'Auria, J. P. Huignard, C. Slezak, and E. Spitz, *Appl. Opt.* **13**, 808 (1974).
165. J. H. Strickler and W. W. Webb, *Adv. Mater.* **5**, 479 (1993).
166. W. K. Swainson and S. D. Kramer, U.S. Patents 4,471,470 and 4,466,080 (1984).
167. S. Esener and P. M. Rentzepis, *Proc. SPIE-Int. Soc. Opt. Eng.* **1499**, 144 (1991).
168. W. Denk, J. H. Strickler, and W. W. Webb, *Science* **248**, 73 (1990).
169. E. Kratzig and co-workers, *Proc. SPIE Int. Soc. Eng.* **164**, 33 (1979); H. Franke, H. G. Festl, and E. Kratzig, *Colloid Polym. Sci.* **262**, 213 (1984); M. Kopietz and co-workers, *Polym. Photochem.* **5**, 109 (1984); M. Kopietz, M. D. Lechner, and D. G. Steinmeier, *Eur. Polym. J.* **20**, 667 (1984).
170. *Chem. Rundschau*, (14) p. 5, (Apr. 1985).
171. M. D. Lechner and co-workers, *Makromol. Kolloid.*, Freiburg, Germany, 1985, *Congress Book*, p. 78; T. Parish, *BYTE* **15**, 283 (1990).
172. M. Irie and K. Sayo, *J. Phys. Chem.* **96**, 7671 (1992).
173. **U.S. Pat. 4,101,976** (1978), G. Castro and co-workers; J. Friedrich and D. Haarer, *Angew. Chem.* **96**, 96 (1984), and *Angew. Chem. Int. Ed. Engl.* **23**, 113 (1984); W. E. Moerner, T. P. Carter, and C. Bräuchle, *Appl. Phys. Lett.* **50**(8), 430 (1987); W. E. Moerner, ed., *Persistent Spectral Hole-Burning*, Science and Applications, Springer, Berlin, 1988; W. E. Moerner, *Jpn. J. Appl. Phys.* **28**, 221 (1989).
174. U. P. Wild, S. E. Bucher, and F. A. Burkhalter, *Appl. Opt.* **24**, 1526 (1985); A. Renn and U. P. Wild, *Appl. Opt.* **26**, 4040 (1987); U. P. Wild and co-workers, *Jpn. J. Appl. Phys.* **26**, 233 (1987); U. P. Wild and A. Renn, *Proc. SPIE Int. Soc. Opt. Eng.* **910**, 61 (1988); Y. Jino and co-workers, *Chem. Phys. Lett.* **140**, 76 (1987).
175. U. P. Wild and co-workers, *Proc. Int. Symp. PME'89*, Tokyo 1989, Kodansha Publ., Tokyo, 1990 and VCH, Weinheim, 1990, p. 507; U. P. Wild, A. Rebane, and A. Renn, *Adv. Mater.* **3**(3), 453 (1981).
176. K. Ao, S. Jahn, L. Kuemmed, K. Weiner, and D. Haarer, *Jpn. J. Appl. Phys.* **31**, 693 (1992).
177. T. P. Carter and co-workers, *J. Phys. Chem.* **91**, 3998 (1987).
178. H. Suzuki and T. Shimada, *Mol. Cryst. Liq. Cryst.* **217**, 165 (1992) and *Jpn. J. Appl. Phys.* **31**, 706 (1992); H. Suzuki and co-workers, *J. Lumin.* **53**, 271 (1992).
179. H. Suzuki, T. Shimada, and H. Hiratsuka, *Chem. Phys. Lett.* **183**, 570 (1991).
180. G. Fuxi, *Proc. SPIE-Int. Soc. Opt. Eng.* **1519**, 530 (1991); D. Wang and co-workers, *Chinese Phys. Letts.* **12**, 556 (1990).
181. H. Suzuki, *Adv. Mater.* **5**, 216 (1993).
182. J. Hennig, *Kunststoffe* **75**, 425 (1985); J. Hennig, *Angew. Makromol. Chem.* **145/146**, 391 (1986).
183. S. Anders and B. Hardt, *Kunststoffe* **77**, 21 (1987); S. Anders, H. Schmid, and K. Sommer, *Kunststoffe* **79**, 55 (1989).
184. **U.S. Pat. 2,938,232** (1960), F. A. Martin (to Hoover Co.); **U.S. Pat. 2,443,286** (1948), Johnson; **U.S. Pat. 4,008,031** (1975), H. P. Weber; **U.S. Pat. 4,828,769** (1989), S. M. Maus and G. J. Galic.
185. A. Iwasawa and N. Funakoshi, in Ref. 11, p. 207; M. Takeshima and N. Funakoshi, *Proc. SID* **25**(3), 219 (1984).
186. G. Kaempf, *J. Inf. Record. Mater.* **18**, 83 (1990).
187. R. M. Pisipati, H. Schmid, and G. Kaempf, *MRS Bulletin* **15**(4), 46 (1990).
188. W. Siebourg and co-workers, *Polym. Eng. Sci.* **30**, 1133 (1990).
189. J. Hennig, *Proc. Int. Symp. on Optical Memory* (1987); *Jpn. J. Appl. Phys.* **26**(4), 9 (1987).
190. G. Kaempf and co-workers, *Proc. Int. Symp. PME'89*, Tokyo, 1989, Kodansha Publ., Tokyo, 1990, and VCH, Weinheim, 1990, p. 549.
191. W. Siebourg, *Kunststoffe* **76**, 917 (1986).
192. D. Freitag, G. Fengler, and L. Morbitzer, *Angew. Chem.* **103**, 917 (1991).
193. Product Information "Makrolon CD-2005", No. 155, 4 89, 5M (1989), Mobay Corp., Pittsburgh, Pa., Anwendungstechn., Inform. ATI 735 (1989), Bayer AG Leverkusen, Germany.
194. H. Guyot, *Plast. Europe*, 289 (July 1992).
195. G. Kaempf, D. Freitag, and G. Fengler, *Kunststoffe* **82**, 385 (1992), and *Kunstst. German Plastics* **82**, 9 (1992).
196. G. Kaempf and co-workers, *J. Polym. Adv. Technol.* **3**, 169 (1992).
197. G. H. W. Buning and co-workers, *Proc. Int. Symp. on Optical Memory ISOM'87*, Poster WC 23.
198. G. Kaempf and co-workers, *Polym. Prepr., Amer. Chem. Soc., Div. Polym. Chem.* **29**, 209 (1989).
199. **Eur. Pat. 249,963** (1987) and **Jpn. Pat. 023,658** (1989), K. Shigematsu, T. Nakagawa, and S. Shuji (to Idemitsu-Kosan Co. Ltd.); **Ger. Pat. 3,835,203** (1990), K. Berg and co-workers (to Bayer AG Leverkusen).
200. **Jpn. Pat. 1,087,621** (1987), I. Takahashi and T. Sugano (to Daicel Chem. Ind.).
201. D. Freitag and co-workers, in J.I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, Vol. 11, John Wiley & Sons, Inc., New York, 1988, p. 648.
202. **Jpn. Pat. 242,474** (1989), (to Daicel Chem. Ind.). **Eur. Pat. 249,963** (1987), K. Shigematsu, T. Nakagawa, and S. Shuji (to Idemitsu-Kosan Co. Ltd.).
203. K. C. Stueben, *J. Polym. Sci.* **A3**, 2309 (1965); **Eur. Pat. 287,887** (1988), L. R. Faler and J. C. Lynch (to General Electric Co.).

204. D. Freitag and U. Westeppe, *Makromol. Chem. Rapid Commun.* **12**, 95 (1991); **Ger. Pat. 3,832,396** (1990), D. Freitag and co-workers (to Bayer AG).

205. G. Kaempf, *Kunststoffe* **82**, 1097 (1992).

206. *Apec-HT Application Technology Information*, ATI 849e, Mobay Corp., Pittsburgh, Pa., 1992; *Anwendungstechn. Information*, ATI 489, Bayer AG, Leverkusen, Germany, 1992.

207. N. Isobe and S. Imai, *Sumitomo Chem. Co. Ltd.*, Chem. Research Lab, 1986.

208. **Jpn. Pat 3,196,612** (1988), T. Kawaki and co-workers, (to Mitsubishi Gas Chem. Ltd.); **Jpn. Pat. 3,199,735** (1987), M. Okamoto and S. Nishiyama (to Idemitsu Kosan Petrochem. Co. Ltd.).

209. V. Serini, D. Freitag, and H. Vornaleken, *Angew. Makromol. Chem.* **55**, 175 (1976); **Ger. Pat. 3,807,098** (1988) and **Ger. Pat. 3,823,196** (1990) V. Serini and co-workers, (to Bayer AG).

210. G. H. W. Buning and R. M. R. Gijssen, *Polym. Prepr. Amer. Chem. Soc. Div. Polym. Chem.* **29**, 211 (1988).

211. Nippon Chemtec. Consult Inc., Research and Development Review Report No. 34, Apr. 1986.

212. H. Jaeger, E. J. Vorenkamp, and G. Challa, *Polym. Comm.* **24**, 290 (1983).

213. B. R. Hahn and J. H. Wendorff, *Polymer* **26**, 1619 (1985).

214. A. Todo and H. Kajiura, *Purasubikekura (Japan Plastics)* **38**, 41 (1987).

215. V. Dragutan, A. T. Balaban, and M. Dimonie, *Olefin Metathesis and Ring-Opening Polymerization of Cyclo-Olefins*, 2nd ed., Wiley-Interscience, New York, 1985, *J. C. Mol. J. Mol. Catal.* **65**, 145 (1991).

216. W. Interthal, lecture, *Polymere in der optischen Informationsverarbeitung*, Frühjahrstagung FA Polymerphysik der DPG, Bayreuth, Germany, 1992.

217. *Mod. Plastics Ind.*, 16 (July 1992).

218. *Hoechst High Chem Mag.* **14**, 48 (1993); *Nachr. Chem. Tech. Lab.* **41**, 724 (1993); *Mod. Plastics Int.*, 11 (July 1993).

219. T. Kohara, M. Ohshima, and T. Natsuume, *Mater. Res. Soc. Symp. Proc., Boston* 1989, paper Q 12.9.

220. **Eur. Pat. 254,184** (1987), W. Kloeker and co-workers (to Bayer AG); **U.S. Pat. 5,052,916** (1991), W. Kloeker and co-workers (to Bayer AG).

221. Chem. Eng. (Sept. 29, 1986); **Jpn. Pat. 58-176,204** (1983), (to Mitsubishi Petrochem. Ind. Ltd.).

222. K. Manabe and co-workers, *Polym. Prepr., Amer. Chem. Soc. Div. Polym. Chem.* **29**, 230 (1988).

223. W. Interthal, Hoechst AG, Frankfurt, private communication.

224. *Optical Memory News* **12**, 42 (1988).

225. H. Ohishi and co-workers, *Polym. Prepr.* **32**, 152 (1991).

226. T. H. Wallmann, M. C. A. Mathur, and M. H. Kryder in Ref. 12, p. 175.

G nther Kampf
University of Aachen

Dieter Mergel
University of Essen

MAGNETIC

Information can be recorded by applying a variety of principles and media. At present most information (~95%) is still stored on paper, 3% on microfiche, and the remaining 2% by magnetic-optical-magneto-optical and semiconductor storage devices. Nevertheless, magnetic recording represents a multibillion dollar industry and is still a growing market.

In this article the main focus is magnetic recording (MR) and rather less on magneto-optic recording (MOR) (see INFORMATION STORAGE MATERIALS, OPTICAL). Both methods are used for professional as well as for consumer applications. The various applications of magnetic recording include audio, video, or data recording. Each application has its own type of media in the form of tape, floppy, or hard disk. At present mor media are only available on hard disks. The trends in magnetic recording technology are continually increasing recording densities and storage capacity with a decreasing price per bit. General references on the subject of recording and related technologies are available (1-9).

Progress to Higher Densities

Since ~1965 the increase in recording density has been by a factor of 1000. Based on this trend, an areal density of more than 300 Gbit in.² is predicted to be available in the twenty-first century (10) for the perpendicular recording mode. This prediction is based on computer simulation studies (11). Since the Intermag Conference of 1992, 10 Gbit in.² longitudinal recordings, having bit areas less than 0.1 μm², have been discussed (12). From a design point of view many improvements can be carried out to realize a recording system having very high bit recording (several gigabit per square inch). Development has shown drastic scale-down of the track pitch, bit-cell length, head gap, medium thickness, and head-medium spacing. Progress in track pitch and bit length for commercially available hard disk systems over time can be seen in Figure 1 (13). The respective necessary contributions from track pitch and bit cell length to realize a 1 Gbit in.² system are not yet commercialized, but have been demonstrated since 1990 (13-15).

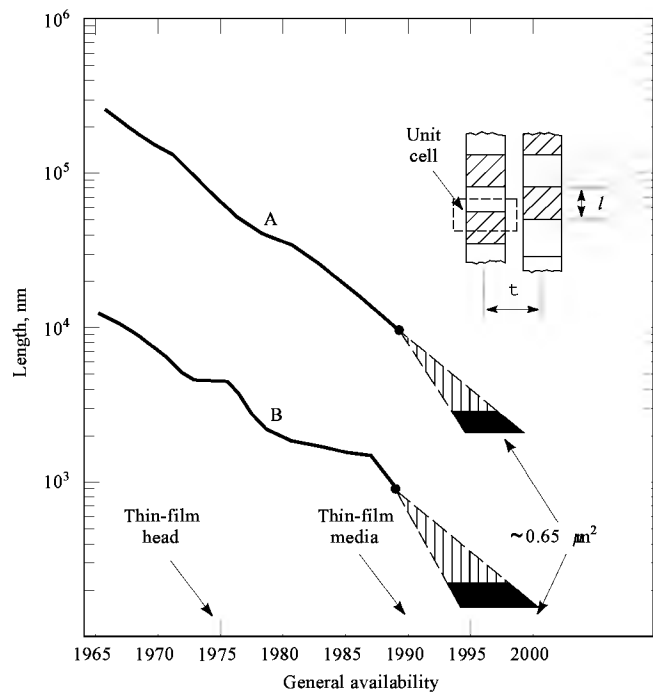


Fig. 1. Progress in track pitch, *t* (A) and bit cell length, *l* (B) implemented in disk files up to 1990. The possible changes in dimensions to obtain commercial 1 Gbit in.² (0.65 μm²) recording are also given (13).

Three key developments have led to increased density: (1) in 1970 a ferrite head was used together with a particulate medium having a coercivity of 28 kA m (352 Oe) with a head-medium spacing of 430 nm, followed in 1980 by (2) a configuration of a plated medium (*H_c* = 56 kA m), a thin-film head, and a spacing of 200 nm. In 1990 (3) a magnetoresistive read head combined with an inductive write head were introduced together with a sputtered medium having a *H_c* = 120 kA m and a head-medium distance of 100 nm. Not only media are being developed with thin-film technologies but also film heads. The small spacing for high densities has introduced an additional field of research on protective layers and tribology.

Thin-film media are available commercially in MO-disks made by sputter technologies, metal evaporated tape (me-tape) for audio application, and electro-deposited and sputtered hard disks for data recording. The next generation of magnetic and magneto-optic recording products will all be dependent on the advances in the volume packing density of recorded information. Consequently there is not only a future for thin-film media but also for thin-film magnetic recording heads (smaller gap width and track width) based on inductive reading and writing as well as magnetoresistive sensors for the read process. Developments of multilayer deposition technologies have been the basis for new media for mor (especially for lower wavelength recording) and giant magneto resistance (gmr) layers for integration in the read head.

The increase of bit densities can easily be demonstrated by showing the bit area (*ba* in μm²), which is determined by the bit cell length (*bl* in μm) × track width (*tw* in μm), for various (commercial) recording systems. In the case of video Hi8 pal (long play) the *ba* = 3.4 μm² (*bl* = 0.2 μm × *tw* = 17 μm), Hi8 ntsc (long play) *ba* = 2.5 μm² (*bl* = 0.25 μm × *tw* = 10 μm), for hdtv (experimental Philips system) *ba* = 2.5 μm² (*bl* = 0.25 μm × *tw* = 10 μm), hdtv (Philips research) *ba* = 1.25 μm² (*bl* = 0.25 μm × *tw* = 5 μm), and digital audio tape (dat) *ba* = 4.2 μm² (*bl* = 0.33 μm × *tw* = 14 μm).

It has been demonstrated by IBM (13-15) that for 1.18 Gbit in.² a hard disk system for longitudinal recording the *ba* = 0.76 μm². Here the bit cell length is 0.19 μm and the track width 4 μm. Hitachi has reported an areal bit density of 2 Gbit in.² with *bl* = 0.21 μm × *tw* = 1.5 μm (16).

With respect to recording systems, modern developments and knowledge of electronics, mechanics, control engineering, etc, have also led to an increase of the densities in actual commercial systems. Thin-film technologies have changed the media and heads, but for increasing the density other

aspects such as the encoding of information are also important and have been changed from analogue to digital. Although digital encoding has already been successfully applied in data recording there is also a trend to use this recording method for video (hdtv) and audio applications (dcc), because errors can be corrected dynamically. Reconstruction of the original information can be carried out very precisely.

Magnetic Properties of Recording Materials

The relation between the three important values for a magnetic material is shown in equation 1:

B = μ0 (H + M)

where *B* is the magnetic induction or flux density generated by a magnetic field *H*. The magnetic induction also consists of a contribution from the magnetization *M*. The *B* in free space is μ₀*H* whereas the contribution from the *M* of the material is μ₀*M*. The vector sum of both is thus *B* = μ₀(*H* + *M*) in which μ₀ is the permeability in free space (μ₀ = 4π × 10⁻⁷ hy/m). The relative permeability in a material is defined as μ_r and given by μ_r = μ/μ₀. A classification for the various types of magnetic materials is given by their susceptibility (χ = *M*/*H*) or permeability (μ = *B*/*H*). Materials are diamagnetic if χ is small and negative (~ 10⁻⁵); another group also having a small but positive χ value (~ 10⁻⁴) are called paramagnetic. The most important materials for recording can be divided into ferromagnetic materials like Fe, Co, Ni (χ = 50 to 10⁴), and ferrimagnetic materials, eg, γ-Fe₂O₃, ferrites. The ferrimagnetic materials also have similar macroscopic behavior to that of the ferromagnets. They both show a spontaneous magnetic moment which means that the magnetization persists even without an applied field. Only these materials have magnetizations large enough for applications. The media used for recording are termed magnetically hard; in comparison with permanent magnet materials it would be better to define them as semihard. The magnetic head materials have the properties of soft magnetic materials. Consequently their properties are very different.

Intrinsic and Extrinsic Properties. The materials Fe, Co, and Ni and their alloys and oxides are mostly used for recording applications materials. Their magnetic properties are described by intrinsic and extrinsic parameters. The intrinsic properties (saturation magnetization, *M_s*, magnetocrystalline anisotropy, *K*, Curie temperature, *T_C*, and the magnetostriction, λ_s) are determined by the type and number of atoms, their arrangement in the crystal structure and their temperature. The extrinsic properties (remanent magnetization *M_r*, coercivity *H_c*, and permeability μ) can also be influenced by the size and shape of the magnetic material and its (magnetic) history. Consequently, in the case of thin-film media, microstructure and morphology play a key role in determining extrinsic properties.

Figure 2 shows the two hysteresis loops for a medium and a head material. The coercivity, *H_c*, the saturation magnetization, *M_s* or induction, *B_s*, remanent magnetization, *M_r* or induction, *B_r*, and the permeability, μ, differ for the two materials.

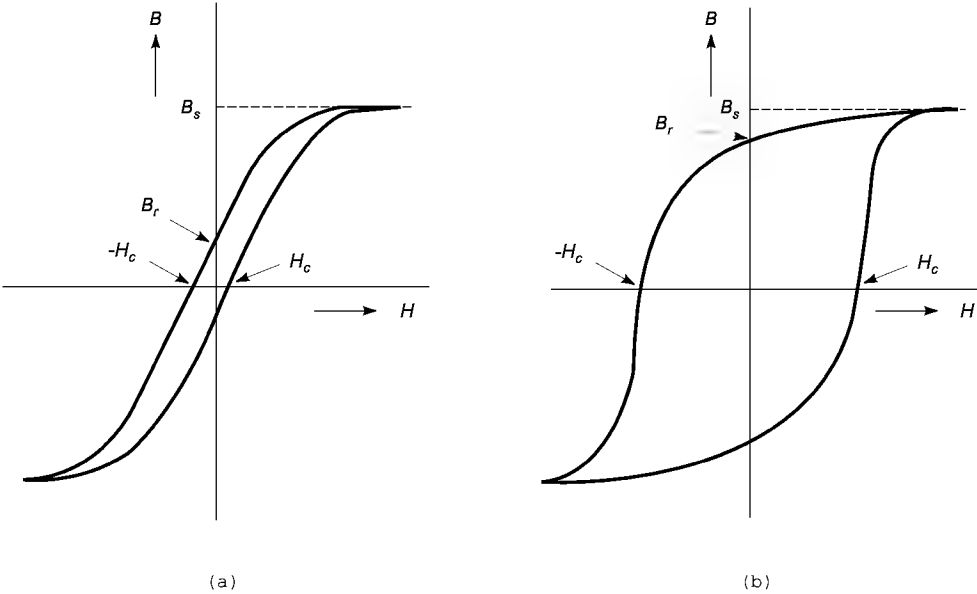


Fig. 2. Basic hysteresis properties for (a) a recording head and for (b) a magnetic medium. See text.

If a very high field is applied the magnetization can reach its saturated state in which all the magnetic dipoles are aligned in the direction of the field. If the magnetic field is switched off, the remanent magnetization *M_r* is left. If the *M* (or *B*) is then reduced to zero, a special field strength, the coercivity, *H_c*, is required.

The hysteresis loop, in general, supplies information about the magnetic properties such as *H_c*, *M_s*, *M_r*, preferred direction of the magnetization or anisotropy, and it can even give some idea about the magnetization reversal process involved. In general, recording media requirements are high coercivity, high remanent magnetization, high squareness (*S* = *M_r*/*M_s*) of the hysteresis loop, and low noise.

The properties for head materials can be summarized as large saturation magnetization for producing a large gap field, high permeability at all frequencies in order to ensure high efficiency, small coercivity with low hysteresis loss, low magnetostriction for obtaining low-medium contact noise, and small but not zero magnetic anisotropy to suppress the domain noise. To ensure good reliability and a long operating time, the head materials must exhibit a good thermal stability and a high resistance to wear and corrosion. The choice of materials and preparation technologies are the tools for tailoring head and medium properties.

The transition temperature at which ferromagnetic behavior of a material changes into paramagnetic is called the Curie temperature (*T_C* for Co, Fe, and Ni, respectively, is 1130, 770, and 358°C). At *T_C* the μ decreases drastically and the *H_c* and *M_r* become zero. The saturation magnetostriction (λ_s) can be defined as the fractional change in length if the sample is saturated from the demagnetized state along the field direction. The λ_s is different for isotropic and anisotropic materials. The field-induced magnetostriction is the variation of λ with *H* or *B* and is an important parameter for magnetic head materials.

Many magnetic materials show preferential directions for the alignment of the magnetization. These directions are energetically favorable and are called easy axes. The energetically unfavorable directions are known as hard axes and are rotated through 90° from the easy axis. When a material has only one easy axis the material is said to have uniaxial magnetic anisotropy. Multiaxial anisotropy occurs in some materials but is less common. The formation of uniaxial anisotropy is not only originated by the crystal symmetry but also by preferential (poly-) crystallite orientation (texture), shape, and size of the sample and stresses.

The strength of the anisotropy determines the difficulty of rotating the magnetization direction away from its stable alignment with the preferred axis and is thus an influencing factor for the magnitude of the coercivity. Anisotropy can also be induced by applying a strong magnetic field during preparation. In the case of multilayers for magneto-optic recording the so-called interface anisotropy (related to the surface anisotropy and strain) is also of

great influence. The shape anisotropy in certain directions depends on the dimensions of the sample and varies with the shape (sphere, elongated particle, thin film). This anisotropy is inherently uniaxial.

Magnetocrystalline anisotropy arises from exchange forces within the crystal lattice and is therefore an intrinsic material parameter in contrast to the shape anisotropy. Materials do have crystal anisotropy if the magnetic moments prefer to lie along special crystallographic axes. This preferential direction of the magnetization leads to a lower energy. These anisotropy direction(s) are different for the various materials used. Easy axes can be found for Fe (bcc): $\langle 100 \rangle$; Ni (fcc): $\langle 111 \rangle$ and Co (hcp): $[1000]$. The hard axes are, respectively, $\langle 111 \rangle$, $\langle 100 \rangle$, and $[1010]$. The magnitude of the first-order crystal anisotropy (K_1) at room temperature for Co is the largest (Co: $10 \times$ Fe). The magnetocrystalline anisotropy is sensitive to temperature and stress and can undergo irreversible changes if the site occupancy of the ions changes. Last but not least the strain anisotropy can play a significant role in the total anisotropy. It can be caused by stresses in magnetostrictive materials. The total anisotropy in a material is the sum of the acting anisotropies.

Switching-Field Distribution. Both M_r and H_c have a strong relation with the recording process. M_r determines the maximum output signal of a recording medium and hence the signal-to-noise ratio. H_c ascertains how easily data can be recorded and erased or changed, but it also determines the maximum head field. On the other hand it also controls the ease with which data can be destroyed, eg, by stray fields. The lower the H_c the more sensitive the medium is to all kinds of fields. In this way, H_c influences the noise level as well. The squareness ratio $S (= M_r / M_s)$ can also be derived from the hysteresis loop. A high value means that a large part of the magnetization is preserved, which is essential for recording.

The slope of the hysteresis loop in H_c is also an important parameter. From this slope, the parameter S^* can be derived (17). In Figure 3 a part of the hysteresis loop (M as a function of the applied field H) is given. The point at which M is constant as the function of the applied field is defined as saturation magnetization (M_s). From the slope at H_c can be written $\tan \theta = M_r / H_c = 1 / (1 - S^*)$ or $dM / dH = M_r / H_c (1 - S^*)$. Thus the S^* is defined in relation to the slope of the loop at H_c . In the case of longitudinal recording experimental data have shown that there is a connection between S^* and recording parameters (18,19). Although S^* is normally used as a switching field distribution (sfd) parameter, it is not always suitable. The sfd can be seen as a distribution function of the number of units reversing at a certain field. For a particulate medium without collective behavior, this function is closely related to the particle size distribution, as differently sized and shaped particles reverse at different fields. Of course the shape, orientation, and interaction between particles influence the sfd as well. Media with a high H_c and a small sfd are more suitable for high density recording (20) because the distribution of the switching fields is very small. An alternative definition is $sfd = \Delta H / H_c$. In this case the ΔH is the full width at half height of the differentiated loop (dM/dH).

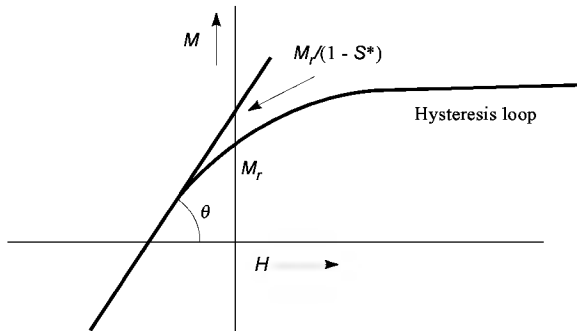


Fig. 3. Williams-Comstock construction for defining S^* .

The micromagnetic structure is directly related to the microstructure and chemical inhomogenities in the layer. The materials used and the deposition technology as well as the parameters play an important role. Thin-film growth, nucleation processes in relation to the deposition parameters, are very important for understanding the thin film microstructure. The relationships between sfd and recording properties are not necessarily valid for media with perpendicular anisotropy as the demagnetizing field can be more important than sfd.

Magnetic Recording

There are two modes of magnetic recording dependent on the direction of the magnetization (magnetic anisotropy) namely longitudinal (LMR) and perpendicular magnetic recording (PMR). In the former the magnetic anisotropy lies in the plane of the medium, and in the case of pmr the anisotropy is directed parallel to the medium normal. Magneto optic recording (MOR) also requires a perpendicular magnetic anisotropy.

Modern recording technologies are based on digital signal processing, even for audio and video recording. Digital technologies provide a much lower signal-to-noise ratio, comfort of error detection, and correction and integration of large-scale integration (LSI) circuits technology. The input data may be either analogue (audio, video; using an a-d converter) or digital (computer data). Examples are the minidisc and dcc (digital compact cassette) for audio application and the not yet commercialized HDTV digital VTR.

Longitudinal Recording. The principles of LMR are given in Figure 4. High density recording depends on shortening the recording wavelengths and also narrowing the track width. The basic recording principles are described in detail in the literature (1–9).

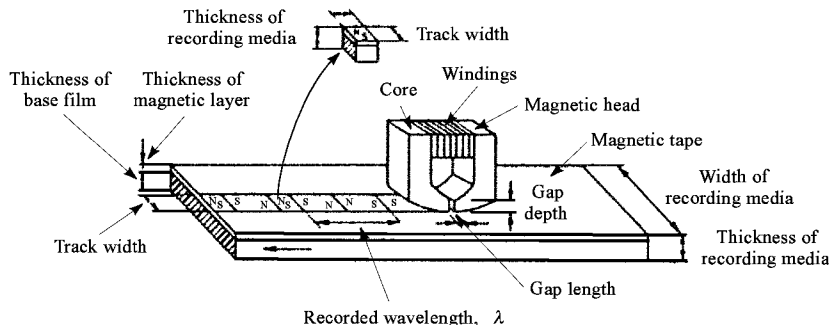


Fig. 4. Written bits in the longitudinal recording mode (LMR). Shortest recorded wavelength is $\lambda_m / 2$.

The digital storage of bits in a moving medium beneath a soft magnetic head can be achieved by a current pulse in the recording head. The magnetic flux associated with the writing current travels the magnetic circuit of the head. The recording medium is in contact with or at a small distance from the gap of the magnetic head. The gap fringe flux penetrates the medium and magnetizes small volumes of the material. In the case of longitudinal recording the

horizontal fringing field component is the most important one. From analytical models and experimental data it is concluded that this field should be several times the value of the coercive field of the medium (1,2,4). The bit size area (bit length $\times$ track width) should be small for high density recording and the transition length (magnetization reversal) as sharp as possible. The change of the head current from one polarity to another produces a magnetization that is not spatially sharp but varies over a distance from one direction to the opposite one. The transition length in longitudinal recording is mainly determined by the head field gradient and the demagnetizing fields generated by the magnetostatic fields of the transition itself and the switching field distribution of the magnetic units in the medium.

A principle configuration of the recording geometry is given in Figure 5. A ring head is used which operates on a head-medium spacing d from the media in motion. The gap length of the head is g , the magnetic thickness δ of the recording medium (in the case of thin hard disk media δ is equal to the film thickness, in the case of thick magnetic particle tape δ is smaller than the coating thickness), and the recorded bit cell size is b with a track width w . A bit cell is defined as an alternative distance between recorded transitions. The bit cell area as recorded has the dimension $w \times b$. The transition length is an important factor for discussing the limitations for the recording density. The transition width a is given by equation 2 where M_r = remanent magnetization, δ = layer thickness, d = head-to-medium spacing, Q = a value related to the field gradient of an inductive head (~ 0.75), and H_c = medium coercivity (17).

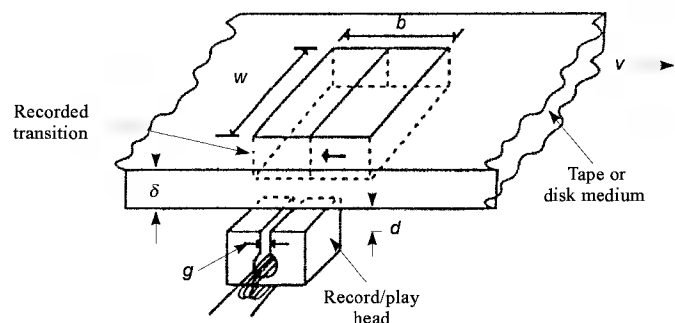


Fig. 5. Basic recording parameters for a medium bit cell and head (7).

$$a \sim \left[\{ 4 M_r \cdot \delta (d + \delta / 2) \} / (Q \cdot H_c) \right]^{\frac{1}{2}}$$

For a small transition width, H_c should be large in relation to $M_r \cdot \delta$ (remanence thickness product) and d as small as possible (contact). The stored bit (transition) can be read by the head due to the flux changes in the head which gives an induced voltage. The output voltage is shaped as in Figure 6. One of the models for the shape of the magnetization transition which is often used is the arctangent model (17) described by equation 3. This model describes the recorded magnetization in the x direction (longitudinal) in response to a step function change in the write head current. The arctangent parameter, a , is inversely proportional to the maximum slope of the transition. The isolated output voltage pulse is given in Figure 6.

$$M(x) = 2 \pi M_r \arctan(x/a)$$

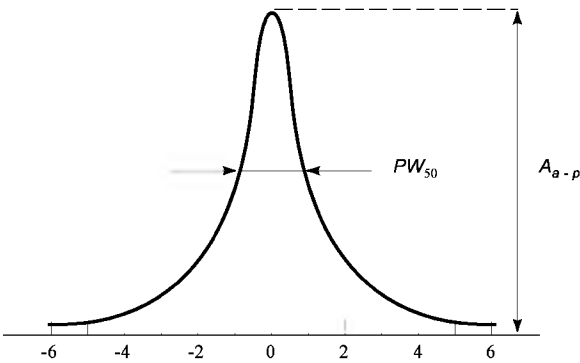


Fig. 6. Output pulse from reading a transition in the longitudinal recording mode.

Narrow transitions (small a) yield large peak voltages (A_{a-p}) and narrow pulse shapes (PW_{50}). The width at half height is given by equation 4 where g = gap length head, d = head-to-medium spacing, a = transition width, and δ = layer thickness (17). A small PW_{50} gives a better resolution without interference from neighboring transitions and a higher read voltage.

$$PW_{50} \sim \left[g^2 + 4(d + a) (d + a + \delta) \right]^{\frac{1}{2}}$$

Medium Noise. In magnetic recording systems three types of noise must be considered: medium, head, and electronic noise. Medium noise is the most important factor influencing the performance of the recording system. Local structural and chemical variations in the medium cause distortions in the magnetization pattern resulting in noise in the signal. The magnitude of noise in particulate medium (CrO_2 , $\gamma\text{-Fe}_2\text{O}_3$) is determined by the properties of the individual particles and the relatively weak magnetostatic interactions between the particles. The packing density (ratio between a magnetic and nonmagnetic material) is the key factor for this noise. In general smaller particles have a statistical reduction of noise. This rule is valid if the medium does not have exchange coupling and small effects of magnetostatic interaction. However, very small particles give a continuous decrease of coercivity until the particles become then superparamagnetic. New materials having higher switching energies have been developed either with higher coercivities or higher magnetization. The density for $\gamma\text{-Fe}_2\text{O}_3$ particles with a particle length of $1 \mu\text{m}$ is about 10_{11} mm^{-3} . Cobalt-modified $\gamma\text{-Fe}_2\text{O}_3$ (length $0.5 \mu\text{m}$) increases density by a factor of 10. The recently developed metallic Fe particles (length $0.25 \mu\text{m}$) have a density of 10^{13} . Smaller particles having higher switching fields (higher H_c and/or M_r) are being developed.

In metal thin-film media consisting of very small grains, a $10 \times$ larger grain-packing density has been achieved but individual grains are not single acting magnetic particles. Depending on microstructural aspects the contribution of the exchange coupling and the magnetostatic interaction must be taken into account. The intergranular distance is only very small compared with that of the particulate medium. Switching of clusters of grains is more obvious.

This results in irregular magnetization patterns in the written transition and therefore to noise in the output signal.

Higher densities and lower noise consequently results from smaller bits containing crystallites with smaller sizes. As an example, a one-bit cell in the Co-based medium used for the IBM 1.19 Gbit in.² demonstration disk (13) consists of about 1900 grains with a grain size of about 20 nm. Consequently, for a 10 Gbit in.² thin-film media a grain size of about 10 nm is necessary (smaller grain size is not possible due to the superparamagnetic limit). In this case there will be about 600 grains in a one-bit area.

Perpendicular Recording. In the case of perpendicular recording the easy axis of the magnetization is perpendicular to the medium (Fig. 7). The most important reason for developing the perpendicular mode of recording is that in the high density area the transition length becomes more and more important. This mode of recording has been studied for application since 1975 by discussing the circular magnetization mode and the technology for preparing a Co-Cr film with a perpendicular anisotropy (21). Since that time Co-Cr has been the most promising material for medium application in perpendicular recording, although other media materials such as Co-O and Ba ferrite films and particles are still under development. Commercial use was announced but not realized in 1989 (22) when Censtor Corp. started working in conjunction with Northern Telecom in the United States. Current strategies about hard-disk contact or quascontact perpendicular recording have been discussed (23). Research activities and published results indicate no specific scientific problems preventing the use of perpendicular recording systems. Perhaps there are commercial reasons for its nonuse such as the continuous improvement of the longitudinal recording technology, eg, the Gbit in.² demonstrations by IBM in the United States (15) and Hitachi in Japan (16).

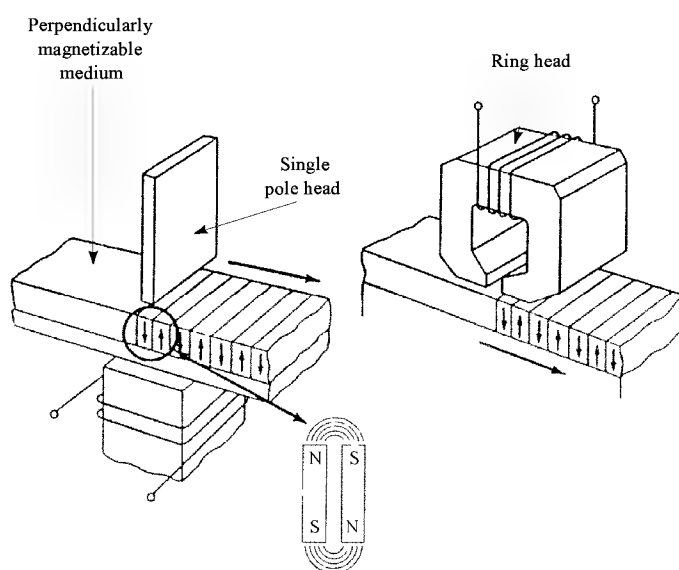


Fig. 7. Principle perpendicular recording for a ring head and a head consisting of a single pole with an auxiliary pole.

Nevertheless, in spite of lack of commercial success the perpendicular mode has stimulated the other modes of recording and presents many new scientific questions that have to be solved concerning new specific heads, contact between head and medium, and the development of suitable thin-film media. The total concept of this type of recording can only be accomplished as a complete success if the media and head development take place as a unified project. For instance, a combination of a perpendicular head and a so-called double-layer medium, as shown in Figure 8, in which the bottom (backing) layer NiFe was prepared as a soft magnetic layer with a top layer of Co-Cr, has been used for very high density recording studies (25).

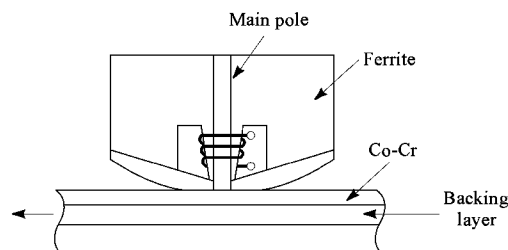


Fig. 8. The main-pole driven head in combination with a Co-Cr/Ni-Fe double-layer medium for perpendicular recording.

A single-layer medium can also be used in combination with a very narrow gap ring head having a high saturation induction core (26). It was clear from the research that the proper high density recording conditions are much more severe for the single-layer medium. Much of the research in the field of perpendicular recording is concentrated in Japan.

Demagnetization. The internal microstructure as well as the macroscopic shape can influence the internal magnetizing field. In the case of PMR and MOR the anisotropy must be perpendicular, which means that the highest demagnetization field is directed opposite to the magnetization of the sample. This means that the anisotropy energy in the perpendicular direction should be larger than the demagnetizing energy for the perpendicular direction of the magnetization.

In Figure 9 the initial states of the magnetization are given for the two recording modes. In both cases two written bits with their directions of magnetization opposite and parallel to the anisotropy direction are drawn. In Figure 9a the transitions with uniform magnetization and zero width can be seen. The corresponding magnetizations and demagnetizing fields are shown in Figure 9b. In contrast to the longitudinal mode, the H_d in the perpendicular mode vanishes at the transition. As a consequence the longitudinal transition will be spread out, in contrast to the perpendicular transition, which is sharpened (see Fig. 9c). In the case of very high density transition recording this principle shows the advantage for the perpendicular mode, but magnetic recording is more than just considering the transition. The recorded bits for lmr media are shown in Figure 4. For very high densities the written bit is enlarged separately in the figure; as the bit length ($\lambda/2$) approaches zero the H_{dr} (see also Fig. 9) in the direction of the media motion approaches one, whereas the transition (M_x) becomes unsharp. At very high bit densities, the longitudinal media is thin (10–50 nm) and the perpendicular media can still be between 50–250 nm. The demagnetizing field also plays an important role, especially in the case where there are media with perpendicular anisotropies. For a thin continuous layer, having no other anisotropy sources, the magnetization prefers to lie parallel to the film plane, being the state of minimum energy. In the case of Co-Cr having perpendicular anisotropy there is, in principle, a competition between the uniaxial anisotropy of a hexagonal structure and the demagnetizing energy of the thin film. In the case of magnetically separated Co-Cr columns (particulate morphology) then also the shape anisotropy contributes to the perpendicular anisotropy.

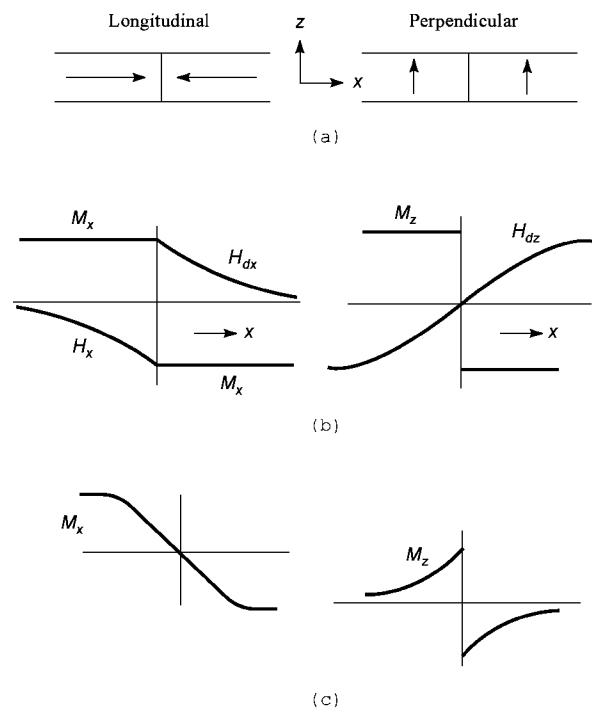


Fig. 9. The relation of an LMR and PMR transition: (a) initial state of magnetization; (b) the magnetization and corresponding demagnetization fields of a; and (c) the resultant magnetization of the different modes.

Most calculations on perpendicular films start at a state of uniform perpendicular magnetization, and from this a so-called quality factor can be derived as $Q = 2K_1 / \mu_0 M_s$. In the case of a uniform magnetized layer with $Q > 1$, the perpendicular axis is thus the direction of the preferred magnetization. However, the state of uniform perpendicular magnetization is quite unusual. Either the magnetization of the layer is split up into domains by the demagnetizing forces, or the layer is in the recorded state. In the latter case, shape anisotropy of a thin film loses its meaning and the demagnetizing energy density (E_d) of a stripe-domain structure (27) must be considered. If the domain width d is smaller than the layer thickness δ , then $E_d = 0.136 \mu_0 d \delta M_s$. The condition for stable perpendicular magnetization ($K_1 > E_d$) expressed in Q becomes $Q > 0.27 d / \delta$. If a Co-Cr layer with typical parameter values is considered ($M_s = 400 \text{ kA/m}$ and $\delta = d = 1 \mu\text{m}$) and $Q > 0.27$. This is much more stringent than in the previous case where $Q > 1$. This simple example shows that a substantial perpendicular magnetization may be expected even if $Q < 1$.

Magneto-optic Recording. Magneto-optic recording is the oldest mode of perpendicular recording; it was demonstrated in the literature in 1957 using a medium MnBi having a perpendicular anisotropy (28). This mode of recording combines the advantages of optical and magnetic recording. The number of commercial systems is growing. The minidisk is a good example for using MO-recording technologies for consumer applications. The status and future of MO-disk drive technologies is discussed in Reference 29.

As can be seen in Figure 10 (30) reversed domains are formed with a focused laser spot (diameter $< 1 \mu\text{m}$) in a perpendicular magnetically saturated disk. Writing is based on the thermomagnetic writing process which means that the magnetization reversal can be achieved by the demagnetizing field or by an applied field, heating the material above its Curie temperature. In threshold writing advantage is taken of the property of some materials that the coercivity decreases strongly with increasing temperature. In this case the magnetization reversal takes place if the applied field plus the demagnetizing field is larger than the H_c during (local) heating.

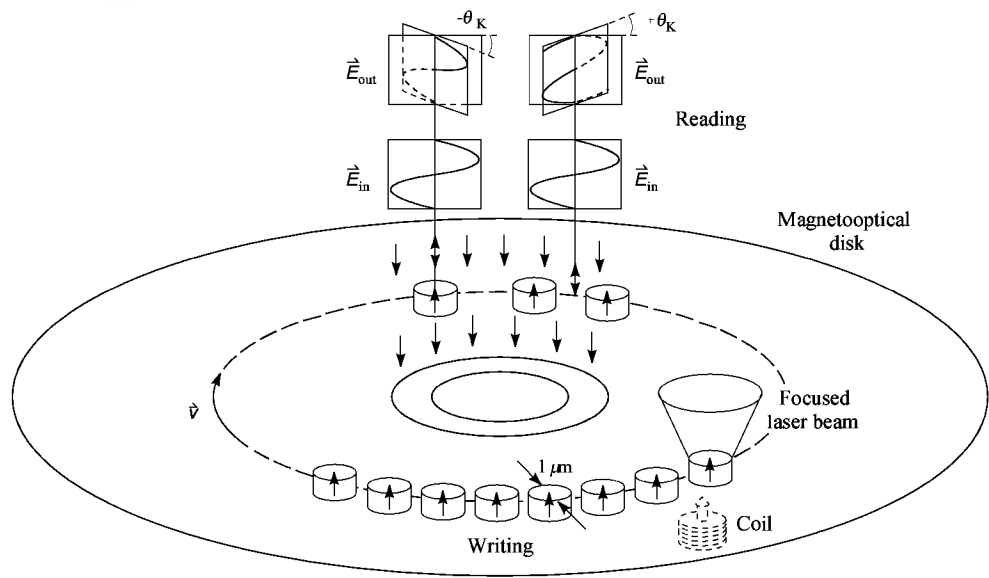


Fig. 10. Written bits in magneto-optic medium (30).

Pulsed diode laser beam has been used for writing at a wavelength of 800 nm and 10 mW/100 ns. The disk is rotated at a frequency of 37 Hz. The external applied field perpendicular to the medium is lower than the H_c of the medium at room temperature in order to limit the switching to the heated spot only. The same laser as for the write operation is used for reading but at a lower power level. Magneto-optical read-out is based on the fact that a change in polarization direction occurs when linearly polarized light is reflected from a magnetic surface (polar Kerr effect). The sign and magnitude of the Kerr angle (θ_K) depend on the direction and magnitude of the magnetization vector with respect to the propagation direction of the light. The polarization vector of the written domains in Figure 10 (magnetization up) are rotated with $+\theta_K$ while the rest of the saturated disk is rotated in the opposite direction ($-\theta_K$).

Erasing of the information can be realized by heating the film with a laser beam while reversing the applied field. This is one of the limitations of mo because it slows down the data rate by a factor of ~2. New media designs (exchange coupled magnetic films) are proposed for realizing direct overwrite by modulating the laser power during the write process (31).

Magnetization Reversal Mechanisms

The reversal of the magnetization is a basic principle of magnetic recording. Magnetization in a material can be reversed by applying a field, and finally the whole material will be saturated in a direction parallel to the field. The two different states of + and - magnetization is the basic idea for digital information storage. The mode of magnetization reversal depends on the material and its size and shape. The two principle methods for reversing the magnetization are rotation and domain-wall motion. Both modes can be examined in terms of energy considerations. The coherent rotation mechanism and the incoherent rotation mechanism only occur in single-domain particles. A particle is single domain below certain dimensions. Above a critical radius a domain wall can exist and the reversal takes place by domain-wall motion.

In the coherent rotation mode the atomic spins remain parallel during the reversal process and may apply only for very small particles. If the particle size increases, incoherent switching mechanisms like curling, buckling, and fanning are used. Magnetic thin films with a polycrystalline structure are strongly exchange coupled and consequently their magnetization reversal takes place by domain-wall motion. In the case of high density recording low noise is a requirement (particulate reversal behavior is necessary for a sharp transition) and certain magnetic properties (high H_c and M_s). By tailoring the microstructural properties, continuous thin films are also promising candidates for actual and future applications.

Switching of a Single-Domain Particle. If an isolated ferromagnetic particle is considered, then the magnetization reversal depends on the dimension of this particle. For recording the so-called single-domain particles dispersed in the medium are most important to deal with. Single-domain behavior can be described by classical nucleation theory (32). A so-called Stoner-Wohlfarth particle is characterized by a uniaxial anisotropy K , the applied field H , and the direction of the magnetization due to H is M . The angle α is between H and the easy axis; θ is the angle between M and H .

The anisotropy energy E_a can be expressed as $K \sin^2 \alpha$. If M is oriented at an angle $(\alpha - \theta)$ with the easy axis this gives a torque of $2K \sin(\alpha - \theta) \cos(\alpha - \theta)$. The torque produced by an applied field, H , is $\mu_0 H M_s \sin \theta$. The equilibrium is given by $2K \sin(\alpha - \theta) \cos(\alpha - \theta) = \mu_0 H M_s \sin \theta$. The M - H loops with the M component in the direction of the applied field for rotation unison are given in Figure 11 for various angles between the long axis of the particle and the direction of the applied field.

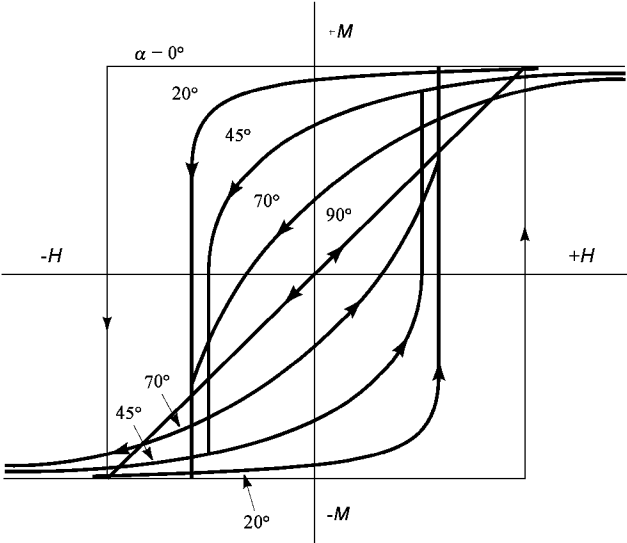


Fig. 11. Magnetization M in the direction of the applied field H for various applied-field angles with the easy axis (33).

At a field H_c at which the particle is saturated perpendicular to the easy axis ($\alpha = 90^\circ$), $H_s = 2K / \mu_0 M_s$. This field can be defined as the anisotropy field H_k . Applying a field in opposite direction in the easy axis, a completely reversible change of M , by pure rotation, occurs without hysteresis. On the contrary, if a field is applied antiparallel to the anisotropy axis the particle switches irreversibly (no rotation) only after applying a field greater than $2K / \mu_0 M_s$.

This critical field called coercivity H_c or switching field H_s , is also equal to H_k . If a field is applied in between 0 and 90° the coercivity varies from maximum to zero. In the case of this special example the applied field $H_a = H_s = H_c = H_k$. Based on the classical theory, Stoner-Wohlfarth (33) considered the rotation unison for noninteracted, randomly oriented, elongated particles. The anisotropic axis can be due to the shape anisotropy (depending on the size and shape of the particle) or to the crystalline anisotropy. In the prolate ellipsoids b is the short axis and a the longest axis. The demagnetizing factors are N_a (in the easy direction) and N_b . The demagnetizing fields can then be calculated by $H_{da} = -N_a M_s$, and $H_{db} = -N_b M_s$. The shape anisotropy field is $H_a = (N_a - N_b) M_s$. Then the switching field $H_s = H_a = (N_a - N_b) M_s$.

If there is a crystal anisotropy, with the easy axis parallel to the shape-anisotropy axis of the particle, the total anisotropy is $H_a = H_{a,shape} + H_{a,crystal}$ and the total switching field is $H_s = (N_a - N_b) M_s + 2K_1 / \mu_0 M_s$ (K_1 = crystal anisotropy constant). In the case of practical materials it became clear that H_c was much lower than predicted by the Stoner-Wohlfarth theory. Therefore the incoherent rotation modes (fanning, curling, buckling) have been explored. The aim of the chain of spheres reversal mode (34) was to give a more realistic picture for an elongated particle used in a particulate recording medium. This reduced the switching field drastically. Furthermore, curling has been introduced as another nonuniform rotation process. In this reversal mode the particle is assumed to be spheroid or an infinite cylinder with its long axis in the field direction. An applied field parallel and opposite to the long direction, the nucleation field, is given by

$$H_n = 2K_1 / \mu_0 + N_a M_s = 2\pi k A / \mu_0 M_s \times 1/R$$

Here K_1 is the intrinsic anisotropy constant due to the crystalline anisotropy. After the demagnetization in the longest direction, k is the shape-dependent constant (for an infinite cylinder $k = 1.38$), A is the exchange constant, and R the particle radius. An infinite cylinder with only shape anisotropy gives $H_n = 6.8 A / \mu_0 M_s \times 1/R$.

Isolated particles, ie, noninteracting particles, are not realistic in recording media. The magnetostatic interaction of the particles should be taken into account. Although the packing densities are not high there is still cluster forming and particles at very close distances. These kinds of effects and also, for instance, the particle-size distributions, influence the switching fields and coercivity behavior.

Below a critical size the particle becomes superparamagnetic; in other words the thermal activation energy kT exceeds the particle anisotropy energy barrier. A typical length of such a particle is smaller than 10 nm and is of course strongly dependent on the material and its shape. The reversal of the magnetization in this type of particle is the result of thermal motion.

Beside single-domain particles, multidomain particles also exist. The distribution in domains lowers the magnetostatic energy but increases the exchange energy caused by the domain walls. The reversal in such particles mainly takes place by domain-wall motion. This kind of reversal mechanism influences the coercivity. Figure 12 gives the relation between the total energy and the particle diameter. The crossing of the curve shows the critical diameter where the particle changes from single- to multidomain. At very low energies superparamagnetic particles can be found.

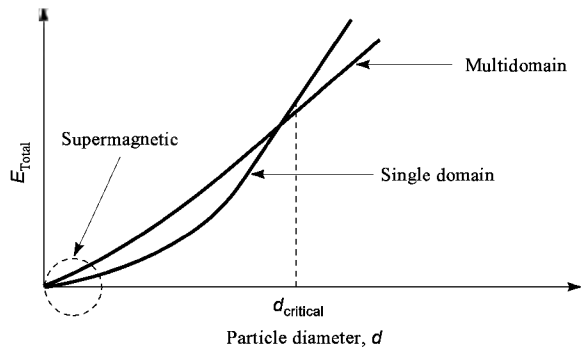


Fig. 12. Total energy vs particle diameter for single- and multidomain particles.

In Figure 13 the relation between the intrinsic coercivity H_c and the particle diameter d is given. The figure is based on a described model (35). The maximum is found around the critical particle diameter. In general the particle diameter and size is not very well defined. For the multidomain particles ($d > d_{critical}$) the H_c is smaller than the intrinsic anisotropy field of the particle. Nucleation effects cause a decrease in H_c as the d increases. This behavior is understood only qualitatively; for a full description see Reference 36. Low noise media should consist of single-domain particles; reversal by domain walls is slow and introduces noise.

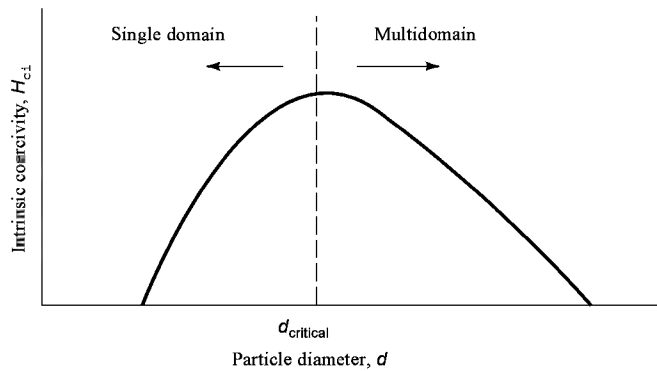


Fig. 13. The intrinsic coercivity H_c vs the particle diameter for single- and multidomain particles (35).

Reversal Mechanism in Thin Films. The ideal magnetic structure for magnetic recording medium consisting of a polycrystalline microstructure is that a crystallite reversed its magnetization by rotation and not by domain-wall motion. In other words, for high density recording the crystallites should act as independent single-domain particles consequently without exchange coupling, but depending on the particle distance still with magnetostatic coupling. In practice the thin films possess a wide distribution of grain size, and not all crystallites are completely separated from each other. This influences the reversal behavior.

There are two main models from the microstructural point of view, namely the particulate and the continuous microstructural model. In the first one the crystals that are formed during film deposition are believed to interact only through magnetostatic interaction. No exchange force acts over the column boundaries due to physical separation. In the continuous model the reversal mechanism is thought to take place by Bloch walls as in stripe domains, hindered by the column boundaries which can increase the coercivity of the medium. This has been studied at length for Co–Cr films having a perpendicular anisotropy (37). Several methods have been used for studying the magnetization process by experiments based on microscopic observations like magneto-optical Kerr microscopy, bitter technique, colloid-sem method, neutron depolarization, Lorentz electron microscopy, and magnetic force microscopy on the one hand, but on the other macroscopic analysis has also been applied by studying the various parameters of the hysteresis loop (measured with a vibrating sample magnetometer) and also the angular dependence of the applied field. Several groups have conducted research in the field of micromagnetic simulations. Studying the intrinsic domain structures can contribute to the understanding of the reversal process involved.

Hysteresis Loop and Reversal Process in Thin-Film Media. Materials used as storage media should have a nonequilibrium behavior which can be designated as a memory. For a magnetic recording medium this behavior is represented by hysteresis loops. The transition between two states of $+M_r$ and $-M_r$ represent the presence of information. Due to the Weiss domain theory the atomic moments in a ferromagnetic material are ordered. The difference between the demagnetized and the magnetized state is due to the dimensions and the number of domains having opposite directions of magnetization.

In the transition from the demagnetized state ($H = 0; M = 0$) to saturation ($H = \text{large}; M = M_s$), small domains (aligned favorably with the field) grow in the direction of the field (wall motion). Increasing the field another reversal mechanism is relevant, namely the rotation of the magnetization into the easy axis. At very high fields the moments lying in the direction of the easy axis, which is close to the applied field, are coherently rotated in the direction of the field. The final state of the material is a single domain (if the applied field is sufficiently high). In this description the magnetization is reversed mainly by domain-wall motion which means movement and bowing of the wall. Bowing of the wall at low fields is a reversible process, but irreversible bowing can occur if the wall is sufficiently deformed. It can be stacked by pinning sites. Many microstructural properties can influence the domain-wall motion such as grain boundaries, defects, chemical inhomogeneities, magnetostriiction, etc. Magnetization reversal by domain-wall motion can also be the origin of very high coercivities.

Such materials cannot be used for low noise recording because for sharp transitions (no zigzag configuration) it is necessary to avoid exchange coupling between the magnetic units (grains, columns, particles) and to lower the magnetostatic interactions. Furthermore, reversal by domain-wall motion, even in a real particulate medium, acts as a noise source. Consequently small nonexchange coupled single-domain grains must be designed. The relation between the applied field and the magnetization for a magnetic recording medium is given by the hysteresis loop. Such a relation is measured macroscopically by a vibrating sample magnetometer and only gives information about the average magnetic properties of the thin film. Parameters like H_p , M_r , S^* , OR, and S can be obtained from the loop and give characteristic information for recording media. Due to the small bit size in a medium it is also interesting to understand the magnetic characteristics on a small scale.

Preparation of Thin-Film Media

The general properties for media are a sufficient magnetization, M , for reading by the head with an acceptable S/N and an acceptable field strength to create a magnetization reversal directly related to the coercivity H_c . The latter parameter should not be too high for successful writing by the head field but it must be large enough to protect the medium against an unwelcome reduction of the signal during storing by demagnetization. For high density recording a significant potential for changing the signal during the required storing time is the self-demagnetizing field originated in the material itself and is proportional to the medium magnetization. Consequently the H_c must become higher for a more strongly magnetizable media and that is also the case if the recording density increases.

Based on the preparation technology and morphology two different types of recording media can be defined, namely, particulate-coated media and thin-film media. The first consists of discrete magnetic particles dispersed in organic resins and the second is created on the substrate (tape, floppy, hard disk) by depositing a continuous layer of a magnetic metal, alloy, or oxide. Although many different configurations have to be discussed for the different type of thin-film media used in the various fields of application, in general the essential design parts of such thin-film media are given in Figure 14. The media consists of a substrate made of glass, aluminum, polyester, etc; a transition (intermediate or seed layer) between the substrate and the magnetic (recording) layer; the magnetic layer(s); and a covering layer. They all consist of different materials, chemical compositions, microstructures, and thicknesses.

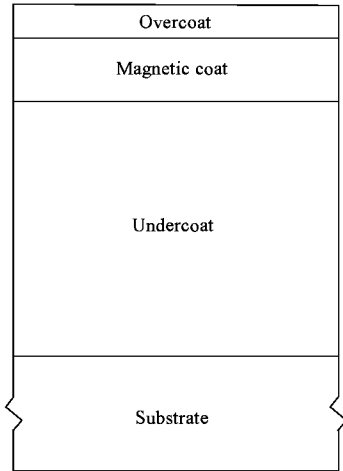


Fig. 14. Typical configuration of a thin-film media.

Thin-Film Media Preparation Technologies. A thin film can be defined as an area (volume) on top of a carrier (substrate) with properties differing from it. The interface between substrate and thin film has a great influence on the properties of the layer. The interface is determined by the properties of the substrate, the material(s) used for the thin film, and the method of deposition. During thin-layer processes the environment can be a liquid, gas, or vacuum. Such layers can be deposited by electro- or by electroless plating, chemical vapor deposition, and physical vapor deposition methods. A close relationship exists between deposition conditions, nucleation, and growth of the layer and their physical properties. Thin layers have properties that differ greatly from those of the bulk materials. These unique properties can be due to (1) their small thickness of a few atomic layers up to micrometer values, which, as a consequence, makes the surface volume ratio of the layer completely different to that of the bulk; (2) because of their typical growth processes they are found in certain microstructures which are, in many cases, directly related to the physical properties; and (3) layer and substrate form a composite system resulting in a combination of properties based partly on the substrate properties and partly on the layer itself. By changing the deposition method and or varying the different deposition parameters various layer structures and morphologies can be created over a wide range. The interaction between layer and substrate, ie, the interface, plays an important role by defining the structure and properties. In contrast to other fabrication methods it is possible to deposit solid materials which can have equilibrium as well as nonequilibrium properties. In the case of thin-film media for magnetic recording four deposition methods have been used, namely, electroless deposition, electrodeposition, vacuum evaporation, and sputtering. More information about deposition technologies and other properties of thin films in general can be found in the literature (38,39).

Using the above-mentioned technologies, multilayers consisting of a few monolayers of ferromagnetic material alternating with a nonferromagnetic material can also be prepared. This technology is used for preparing media for magnetooptic recording and thin-film heads based on the magnetoresistance principle. The most important technologies used for preparing thin-film recording media will be discussed. Electrodeposition and autocatalytic plating (electroless deposition) were initially investigated in the early 1950s. The first magnetic disk for digital magnetic recording was introduced in 1960 and made by electrodeposition. Most of the media (ca 1993) are prepared by physical vapor deposition technologies such as evaporation and sputtering.

Vacuum Evaporation: Oblique-Incidence Deposition. Evaporation processes are usually carried out under vacuum within a pressure range of about 10^{-3} to 10^{-7} Pa (10^{-5} – 10^{-9} torr). The various steps in the production of thin films with vacuum evaporation can generally be subdivided into the creation of the vapor-phase species, transport from source to substrate, and nucleation and growth on the substrate. The material flux is produced by the evaporation source which heats the material to the sublimation temperature: this can be done by resistance, radiation, eddy currents, electron and laser beams, etc. After evaporation the flux condenses on a cooler substrate. The low pressure is essential for having as few collisions as possible with the background gas species (a straight-line path) and a clean process. The emission characteristics of the source are discussed in detail in the literature. The growth speed on the substrate is not equal to the evaporation rate of the source and depends on the deposition geometry, the emission characteristics of the source, and the condensation coefficient in turn depending on the surface conditions and substrate temperature. Oxides, nitrides, etc, can be prepared by adding reactive gas during evaporation.

Because this method is important in relation to the preparation method of metal evaporated tape it will be discussed in more detail. The incidence angle of the atomic flux plays an important role in the nucleation and growth process. Atoms arriving under an angle at the substrate usually show a different behavior to those which approach perpendicularly. First of all, the shadowing effect plays an important role in the film formation. If there is no adatom (physically adsorbed atom) mobility and the sticking coefficient probability equals one, an incoming atom is captured as soon as it touches the substrate or surface atom. Atoms already deposited and surface irregularities throw a shadow. No direct impingement is possible in this shadowed area. When nonzero adatom mobility occurs, which is normally the case due to the kinetic energy of the atoms and the substrate temperature, they can then move to energetically favorable positions including the shadowed areas. Therefore, an increasing mobility partly annuls the shadowing effect. The degrees of shadowing and mobility are determined by the deposition conditions. The direction of the adatom movement is considered to have two contributors, namely surface diffusion (responsible for the movement in all directions), and the angle-of-incidence effect which causes the atoms to have a momentum component parallel to the incidence plane. This is true for the situation where the substrate is fixed in relation to the incoming flux.

The first paper about NiFe layers evaporated under an angle was published in the early 1960s (40). The films prepared this way are often called oblique-incidence or angle-of-incidence films. It was found that these kinds of films show an anisotropy whose strength depends on the angle of incidence of the atoms (α_i) during deposition (Fig. 15). If α_i is between 0 and 65 the anisotropy lies parallel to the film plane and perpendicular to the incidence plane.

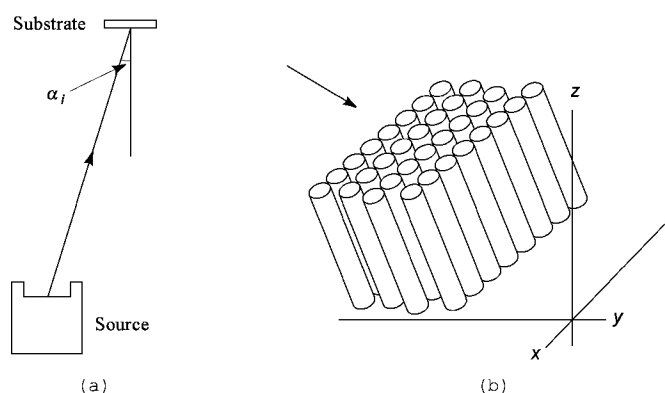


Fig. 15. Oblique-incidence evaporation (a) and a possible columnar structure (b).

Figure 15 gives an impression of the incidence angle of the evaporated flux and the final columnar structure. There is an anisotropic column aggregation perpendicular to the evaporation plane (y - z -plane). The shadowing effect causes the columns to be separated by open regions. In the transverse direction (x -direction) there is a homogeneous distributed arrival of the atoms. The result is a higher packing density in the x - than in the y -direction. An explanation for the anisotropy based on the self-shadowing mechanism which means that a growing cluster of adatoms causes the shadowing of an adjacent region with respect to the vapor source has been given (40). Consequently, the film grows in chains (columns) oriented perpendicular to the plane of incidence. The shape anisotropy could be the reason for the macroscopic anisotropy. From a study of the effect of the mobility of metal atoms on the structure of oblique-incidence films the mechanism of inhibited mobility was introduced (41). With this principle it is possible to give an explanation of the dependence of the preferred direction of growth of texture, angle of incidence, affinity of oxygen, substrate and source temperatures, melting point of the evaporated material, and background gas pressure. Other papers (42,43) have stated that the shadowing mechanism alone is sufficient to explain the majority of the features of the microstructure. However, the modeling work in the papers used an atomic relaxation term which shows a strong resemblance to the mechanism of inhibited mobility. Most of the publications dealing with oblique incidence magnetic films have been produced by Japanese researchers, for example (44).

The relation between the columnar inclination angle (β) and the angle of incidence (α_i) of the evaporation flux was found to be $2\tan\beta = \tan\alpha_i$ (45). The so-called tangent rule is not always valid in a variety of experimental situations.

Besides columnar growth there is another aspect of the morphology, namely columnar bundling (46). Bundling is the growth of a column in the direction perpendicular to the evaporation plane. Bundling is also dependent on α_i , substrate temperature, deposition rate, and gas pressure, and bundling is even found in the direction of the evaporation plane. The interest in oblique-incidence thin films for magnetic recording purposes is mainly created by the experimental results that enable the magnitude and direction of the magnetic anisotropy, the coercivity, and a suitable squareness to be varied by modifying α_i . An overview paper has been published dealing with the oblique-incidence method especially when applied to perpendicular anisotropic thin films. The method consists of two opposite oblique sources which also makes it possible to tailor the chemical composition at the columnar boundaries (47).

Sputtering. From the physical point of view sputtering is a different process from vacuum evaporation. Generally the sputter deposition process implies the ejection of atoms from a target by energetic particles. The ejected atoms then condense on the substrate to form a thin film. The accepted theory of sputtering is based on a momentum transfer process. Therefore the sputtered atoms leave the target (source) with an appreciable kinetic energy (3–10 eV). A part of this energy is dissipated by collision processes with atoms of the sputtering gas. Upon arrival at the substrate the energy is still 1–3 eV whereas for vacuum evaporation it is smaller than 0.1 eV. Typical deposition rates are 0.5–50 nm s⁻¹.

There are two principle sputtering methods, namely glow discharge sputtering and ion-beam sputtering. In the case of glow discharge sputtering (normally called sputtering) a glow discharge is formed between a negative anode (target) and the cathode (substrate holder) at earth potential. The anode potential is on the order of keVs. In the deposition chamber a so-called sputter gas (Ar, Xe, Kr) is admitted up to a pressure of about 1 Pa (10⁻² torr) depending on the system and the type of sputtering.

The plasma is a partially ionized gas composed of electrons, ions, and a variety of neutral species. In principle the plasma is electrically neutral and the particles are uncontrollable. Interaction between the plasma particles and the substrate film surface influences the growth. More information about sputtering can be found in the literature (48). Although there are many parameters which influence the final results of the layer such as the sputtering mode, the geometrical arrangements in the deposition unit (target-substrate distance, target size, substrate size), the type of targets (alloyed or multitarget), the most important sputter parameters are substrate temperature, argon pressure, and power. An overview of sputtering parameters, structural aspects, and the magnetic behavior of Co-Cr films for perpendicular recording is given (49).

Multilayer Technologies. During the 1970s and 1980s enormous advances in thin-film preparation processing have been made in the field of so-called artificial structuring of materials; semiconductors, metals, and insulators have been prepared in various sizes and geometries. Several classes of layered structures have been made using metal compounds. Figure 16 shows a schematic representation of the various possibilities of obtaining structurally ordered multilayers. Generally, the total multilayer thickness (Fig. 16a) is in the tens of micrometer range whereas the individual layers (Fig. 16b) can be varied from one to tens of nm. Layered structures can be deposited by sequential depositing of two or more materials. After preparing a multilayered (metal) film, by alternating deposition of two elements, a periodicity along the film normal should appear if the following conditions are satisfied: the layer thicknesses are determined on an atomic scale, a layered structure is formed, and the interdiffusion is sufficiently suppressed.

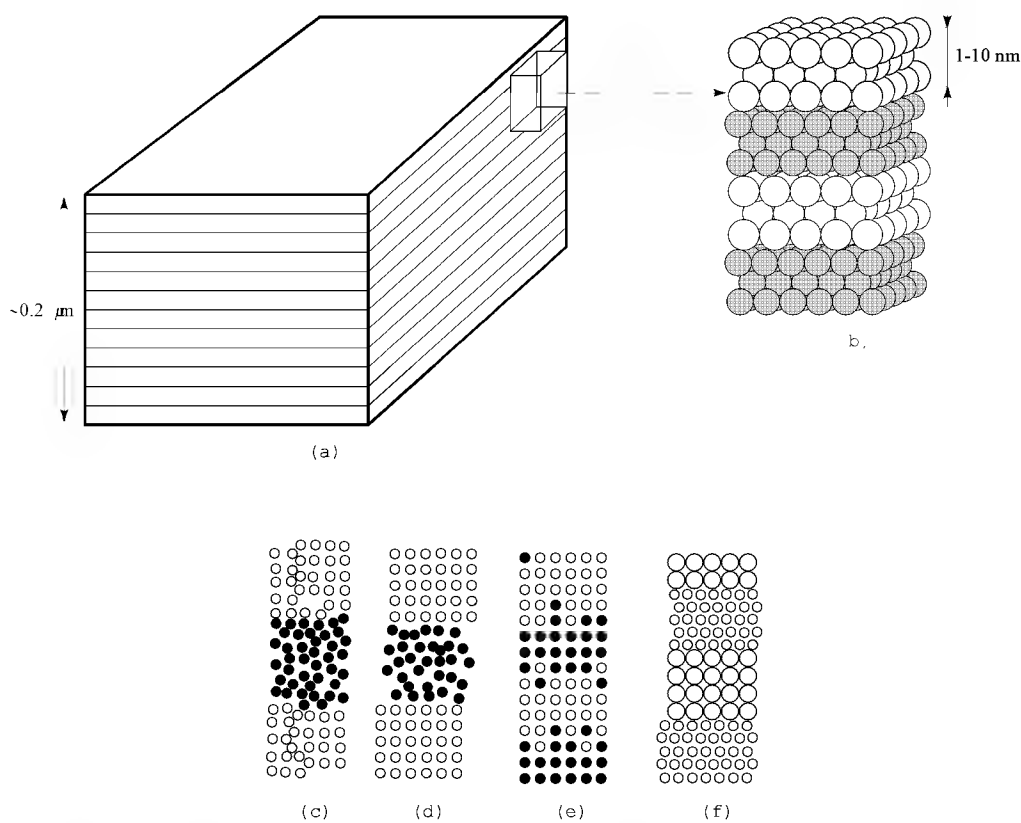


Fig. 16. Schematic presentations of the multilayer structure (a) and (b) and some different possibilities for stacking the individual thin layers (c–f).

The choice of materials for metallic systems is still expanding and at present various examples of combinations with different atomic radii are being prepared. Here multilayered techniques also show possibilities for new material syntheses. In contrast to materials prepared by chemical procedures, superlattices are made far from equilibrium. The various possibilities for layering the artificial superlattice materials are given in Figure 16c–f. Most of the stacked layers (c, d, f) have more or less sharply defined boundaries and some have a noncrystalline structure in the individual layers (c) or one of the layers is noncrystalline (d). In such situations the structural information is not transmitted between adjacent layers and therefore, strictly speaking, no superlattice is formed. In the case of an unsharp boundary (e), compositionally modulated alloy-layered structures have been made. The amplitude of composition modulation in the center of a layer can be in the range of 0 to 100%. Superlattices can also be formed with sharp boundaries (<5% of the thinnest layer) between the two components.

All these possible combinations strongly affect the electronic and physical properties of materials, especially when the dimensions of the layer structure become comparable with the characteristic lengths relevant to the properties in the particular materials. The application of multilayer structures in recording technologies plays an important role. Multilayer configurations are used in thin-film heads (magnetoresistance type) for magnetic recording as well as in media for magnetooptic recording. Crucial aspects important for these types of layers are the sharpness of the interface and the flatness. These aspects are strongly related to the method of preparation, the type of materials, and the substrates used. Generally speaking the layer growth mechanism plays the key role in the final interface structures.

Microstructure and Morphology of Thin-Film Media

The process parameters (flux rate, substrate temperature, etc), type of material (desorption, dissociation, and diffusion-energy terms), and the substrate properties influence the growth process. Depending on the process, film materials, and substrate behavior, all types of layer structures can be grown (amorphous, polycrystalline, and single crystal). Higher mobility of adatoms makes it possible to create films with deviating stoichiometry. Variations in substrate temperatures makes it even possible to deposit metastable structures. With treatment of the substrate surface or deposition of prenucleation centers it is also possible to grow films with a preferential crystallographic orientation (texture) and a specific morphology. A thin-film microstructure can be modified by means of substrate temperature, surface diffusion of the atoms, bombardment during film formation, incorporation of impurity atoms, and the angle-of-incidence effect of the incoming particle flux.

In most cases the final properties of the deposited layers differ principally from materials made by standard metallurgical methods. The substrate temperature is the most important process parameter for explaining the morphology of evaporated films (50). This result is modified for the sputtering process (51) and is extended with a second parameter, namely the argon pressure. In this work the influence of the surface roughness is also considerable. An impression of the structural features for sputtered films is given in Figure 17 called the Thornton diagram.

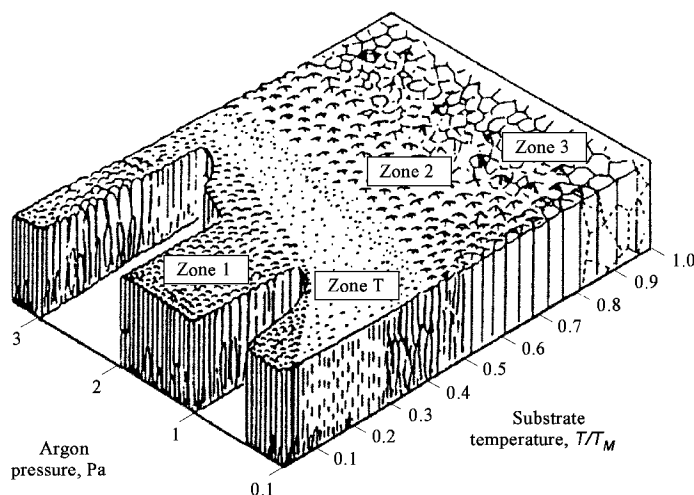


Fig. 17. Structural diagram (51) for sputtered layers. Zone 1 is a porous structure consisting of tapered crystallites separated by voids, Zone 2 shows columnar grains, and Zone 3 has a recrystallized grain structure. Zone T is a transition structure consisting of densely packed fibrous grains.

This zone-structure model has been revised (52) and accounts for the evolutionary growth stages of structure developments as well as the separate effects of thermally and bombardment-induced mobility. The reported experimental results between the microstructural aspects and deposition parameters should be used very carefully for personal results. Variation of the experimental parameters, using different types of materials, can result in divergent behavior. Generally, deposited thin films have higher defect densities than those of bulk materials. The defects in polycrystalline thin films are grain boundaries, column boundaries, voids, vacancies, dislocations, and interior gas bubbles. Defects are mostly responsible for the low temperature and interdiffusion processes. In the case of polycrystalline films the grain boundary is the most important detail. Epitaxial growth is a very special form of nucleation and growth and has a unique orientation relation with the substrate. Single-crystalline films can be prepared by the correct choice of the substrate material and deposition parameters.

Columnar Structure and Grain Size. Most of the deposited films reported in the literature have a so-called columnar structure (see Fig. 15). Depending on the substrate materials the columnar diameter can increase with the layer thickness or it is constant through the layer thickness. Further, the columnar diameter depends on the argon gas pressure during sputtering, the substrate temperature, and the bias voltage on the substrate. The method of deposition also has a large influence on the development of the columnar growth. In the case of Co-alloys deposited on polymer substrates extraordinary column or nodular growth is found. It can be seen from tem observations that the average crystal size has a 10–50 nm diameter. Depending on the preparation conditions the crystals can be mechanically separated (spacing about 4 nm) at the grain boundaries. Another separation of the crystals columns is by segregated crystal boundaries with a nonferromagnetic material. This has been shown (53) in CoP plated disks. Grain size or columnar size are strongly influenced by the microstructure of the underlayer substrate.

Crystal Structure and Texture. Many of the Co–X–Y thin-film media where X(=Cr, P) and Y(=Ta, Pt, Ni) do have a hexagonal close packed (hcp) structure with the texture axis (c -axis) parallel or perpendicular to the substrate surface depending on the properties of the Cr underlayer. It is also possible for the films to have fcc phases. The texture of a polycrystalline material can be simply defined as the crystallographic preferential orientation. In the case of a ferromagnetic thin film the orientation distribution of the column is of great importance because it determines the quality of the direction of the magnetic anisotropy. The texture axis should be perpendicular or in-plane oriented depending on the recording mode.

Another influence on the orientation is caused by the pressure of sputter gas and a worsening background pressure. It has been shown that a small amount of residual N_2 gas in the process chamber causes the formation of the fcc phase which also destroys the well-oriented hcp structure. Usually a higher Ar pressure means that the atoms are more scattered during their movement from target to substrate and consequently their kinetic energy is reduced. Therefore, for producing well-oriented ($\Delta\theta_{50} < 5$) films sputter equipment has to be optimized for certain process parameters. Data has been published on deposition of nonmagnetic underlayers or seed layers to improve the nucleation and growth of the Co–X–Y layers and increase the c -axis orientation.

Chemical Inhomogeneities or Compositional Separation. Compositional separation at the grain boundaries influences the magnetic interactions of the individual grains. Deposition parameters such as temperature, substrate material, and the use of a seed layer play an important role. There are, in principle, two driving forces for obtaining the compositional separation, namely the temperature and deposition geometry.

Detailed studies of the chemical inhomogeneities in thin films for recording applications are seldom found. Most of the work that has been done has focused on Co–Cr media for perpendicular recording. Chemical separation of Cr in Co is very often called segregation, and this can be found at the column boundaries or even in the column itself. The boundary segregation is also described as oxygen gettering (54) which means that during the growth the Cr mainly reacts with oxygen; consequently the M_i of the magnetic composition increases. Another explanation is that there is a recombination of atoms, molecules, or clusters before they interact with the substrate (55). The formation of Cr–Cr and or Co–Co clusters produces a compositional change. The same result can be obtained by using nonhomogenous targets (in the case of sputtering). Cr concentration on the lattice faults present in the column have been mentioned (56). Other explanations, where the Cr migrates to the boundary for thermodynamic reasons, can be found (57–59). These explanations are based on the fact that the Co–Cr system strives to attain a low as possible surface energy by the enrichment of the boundaries with Cr atoms because Cr has a larger surface than Co and also the binding energy is lower. The Cr atoms are exchanged with Co atoms at the surface and also at the column boundaries. The Cr distribution is dependent on deposition parameters such as temperature and energy of the incoming particles.

Often the phase diagrams of bulk Co–Cr systems have been used to explain the chemical composition although such a diagram is only valid at thermodynamic balance. A complete overview and new data have been published (60). With respect to the application of Co–Cr as a thin-film media for magnetic recording the most interesting area of the phase diagram is around the temperature range from room temperature to 1000°C. It can be seen from this phase diagram that at 35 at. % Cr and lower two hcp phases, one with a high Co-rich composition (ferromagnetic) and the other with a high Cr-rich content (paramagnetic) can exist. These phase diagrams are based on an equilibrium process and achieved from thermodynamic calculations. The thin-film materials discussed here are made by deposition which is by definition a nonequilibrium process. In studying these an important fact was assumed, namely that the phase diagram at high Co concentration below 800°C is very complicated because sluggish diffusion inhibits the equilibrium. It is known, for instance, that during sputtering the surface temperature of a grown film is quite different from that of the substrate (48). The bombardment exerted by various particles from the plasma on the surface results in a much higher temperature at the surface.

Most studies of the magnetic behavior of Co–Cr thin films have suggested that there is Cr segregation at the columnar boundaries which can explain the higher coercivity and the magnetic reversal behavior of the layers. However, only a few research groups have experimentally shown the existence of such a compositional separation. Nevertheless, the grain boundaries are assumed to be fast diffusion paths at which diffusion takes place by vacancies (61). Besides the segregation at the boundaries, Maeda (62) first introduced the so-called chrysanthemum-like pattern (CP model). This is also related to the thermal driving force. In this model the enhancement of Cr at the boundaries is a special case of the CP pattern (Fig. 18).

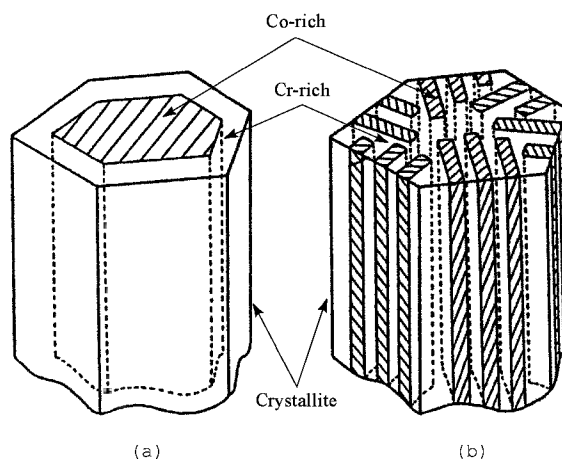


Fig. 18. Conventional segregated columns (a) and the CP structure (b).

The CP structures are observed by tem but also in the case of very thick samples by sem (63). These kinds of observations are only possible in combination with selective etching techniques (64). Concentration fluctuations in the grains (columns) have also been studied by atom probe (65). These types of measurements can only be made after a special preparation technology of the Co–Cr film and substrate. The results show the atom-probe concentration depth in the planar direction of a Co–Cr column with an average composition of 22 at. % Cr. Over a depth of about 40 nm compositional differences of 30 at. % Cr and 7 at. % Cr have been measured. The latter composition is ferromagnetic whereas the composition above 26 at. % Cr is paramagnetic. The fluctuations are less than 10 nm. The main features of the CP structure are Co-rich stripes that tend to be perpendicular to the column boundaries. The CP structure also developed as a function of the thickness of the layer but mainly as a function of the substrate temperature. The Co-rich and the Cr-rich components strongly influence the magnetic properties. The spatial periodicity of compositional separation in sputtered Co–Cr layers on top of soft magnetic Ni–Fe layers is in the range of 3–7 nm (66). It has also been shown that compositional separation is present in Co–Cr–X materials used for longitudinal recording (67). Consistent results from experimental studies of field ion microscopy (fim), nuclear magnetic resonance (nmr), and transmission electron microscope (tem) in combination with etching techniques are presented.

Co Binary and Ternary Alloyed Thin Films. Most of the thin-film media for longitudinal and perpendicular recording consist of Co–X–Y binary or ternary alloys. In most cases Co–Cr is used for perpendicular recording while for the high density longitudinal media Co–Cr–X is used ($X = \text{Pt, Ta, Ni}$). For the latter it is essential to deposit this alloy on a Cr underlayer in order to obtain the necessary in-plane orientation. A second element combined with Co has important consequences for the Curie temperature (T_C) of the alloy, at which the spontaneous magnetization disappears. The T_C for Co is very high, 1131°C, and provides a reasonable margin for the amount of an additional element by which T_C usually decreases. It is also true, at least for bulk material, that adding Cr to Co can finally make the hcp structure become unstable and at certain compositions it is also possible that two crystal structures (fcc and hcp) are present. Furthermore it has been shown that even two different phases, one with a high Co-rich composition (ferromagnetic) and the other with a high Cr-rich content (paramagnetic) can be formed (68). Consequently, adding Cr to Co has two important effects; reduction of the T_C and M_s . For recording applications these values should be optimized. The T_C must not be too close to room temperature, because then the magnetic behavior becomes too sensitive for temperature variations. M_s should have a certain value because otherwise the information cannot be read by the head. The physics behind the reduction of M_s and T_C are complicated and not completely known. However, the most useful model when Cr is added to Co is to consider that the magnetic moment of Co atoms is reduced by electron transfer to its 3d band from Cr. It has been shown experimentally that the T_C drastically decreases with the Cr content and becomes paramagnetic just above about 22 wt % at room temperature. This is not expected if Cr only acts as a simple dilutant (69). Also, the transfer of 4s electrons from Cr to the 3d shell of Co may lower the magnetic moment (70). Furthermore, pure Cr is antiferromagnetic at room temperature and a ferromagnetic sublattice coupling also seems to be an acceptable explanation for the relative strong decrease of T_C when compared with other X elements, which form an hcp phase with Co (71). Adding Cr to Co also gives an enhanced anisotropy field. Furthermore, with the variation of the Cr content it is possible to adjust H_c .

Another favorable influence on the film morphology is the reduction of the column dimensions (54) and the appearance of the compositional separation. The latter has a great influence on the magnetic microstructure because it can lead to more or less magnetically uncoupled columns if the enhancement of Cr at the column boundaries becomes higher than about 27 at. %. Smaller grain size has the advantage of a lesser surface roughness which results in a better head–medium interface. The choice of Cr also improves the corrosion resistance and mechanical hardness of the Co-based medium, although in a few cases other elements have been added to increase these properties. Also quaternary alloys Co–Cr–Pt–X ($X = \text{Ta, B, Ni, B}$) have been reported (72,73). The basic magnetic properties such as magnetization, coercivity, and anisotropy depend on the microstructural properties, compositional separation, and phase separation.

Surface Properties. In the case of very high density recording the surface becomes more and more important. On the one hand the surface smoothness and wearability are important because of the fact that the head–medium distance is very close and on the other hand for writing as well as reading the magnetic behavior is the key factor (74–77). Therefore, analyses of the chemical and structural properties (eg, surface topology) in relation to the magnetic properties are necessary. One conclusion is obvious, namely that for films with different surface and bulk hystereses, the magnetization cannot be homogeneous throughout the film thickness during all stages of the hysteresis curve. Therefore, this aspect should be studied in more detail because as film media are becoming still thinner, the surface volume ratio will be more important.

Microstructure and Magnetic Properties of Thin-Film Media

Magnetic Structure. An important characteristic of a medium is its magnetic structure, the magnetic unit (intrinsic domain structure or written bit) in the magnetizable layer which has, in principle, two oppositely stable directions parallel to the anisotropy axis. The switching of the magnetic units can be achieved by a sufficient applied field. Study of the magnetic structures and their switching behavior, etc, can be carried out by several techniques. The study of the M – H loop gives information about the macroscopic behavior of the media. Increasing density requires more knowledge about the micromagnetic behavior. More insight can be obtained by computer simulations (7). New experimental methods are available and being developed for collecting more information about the mesomagnetic (an area between macro and micro) properties of the media like the methods for observing the magnetic domains, domain walls, written bits, and stray fields using the Bitter-colloid sem method, magnetooptic Kerr (MO-Kerr) observations, Lorentz tem observations, electron holography, scanning electron microscope polarization analysis (sempa), anomalous hall measurements in combination with photolithography, and the magnetic force microscope (mfm).

The microstructure of the thin-film medium has a great influence on the magnetic behavior of the film. Is the magnetic behavior of the layer continuous or particulate? Both qualifications refer to the degree to which exchange forces are able to extend throughout the medium. In a continuous medium the exchange forces are hardly disturbed by structural discontinuities such as crystal boundaries. As a consequence the magnetic-domain boundaries then usually consist of Bloch line walls, which contain exchange and anisotropy forces. The typification particulate refers in the first instance to the method of preparation, whereby particles, usually single-domain particles, are compounded together with nonmagnetic materials. In those media only

the exchange forces are restricted to the volume of the particles, which therefore show only just magnetostatic interactions. Within this definition, from the magnetic point of view, it is possible that even thin films can be considered as particulate, showing distinct structural ferromagnetic units like crystals, columns, or clusters separated by nonferromagnetic materials or voids. Such kinds of microstructures can be influenced by the deposition methods and the nucleation and growth process of the layers. If the thin film has a continuous microstructure in which there is an exchange between the magnetic units, then at a remanent magnetic state (B_r in Fig. 2) the layer consists of magnetic domains having their direction of magnetization antiparallel and separated by a domain wall. The wall is a transition region in which the spins are rotated from one direction (domain A) in the other direction of domain B. The thickness of such a 180° wall is determined by minimizing the various energies and is, of course, dependent on the type of material (Co = 8.4 nm, Fe 30 nm, and Ni 72 nm).

The magnetization thin-film medium is strongly dependent on the microstructural properties. Depending on the morphology and chemical inhomogeneties of thin-film media the reversal take place as follows. Exchange coupled grains show a magnetic-domain structure which covers many grains and gives a reversal originated from domain-wall motion. The wall energy, influenced by crystalline anisotropy, magnetostatic energy at the wall, strain, chemical inhomogeneties, and film surface properties, dominates the coercivity H_c . Uncoupled grains which when reversed independently are, of course, influenced by magnetostatic interactions. Here the H_c is determined by crystalline anisotropy of the grain, shape, and strain anisotropy. Clusters of grains which are locally magnetically coupled can reverse in unison. These reversals are independent of other clusters, but again magnetostatic coupling can influence their behavior.

The shape of a magnetic material (sample geometry) is the most obvious feature which may influence the anisotropy. Depending on the geometry there is a magnetic charge at the surface of the uniformly magnetized sheet, cylinder, ball, or sphere. This magnetic pole density produces an internal uniform demagnetizing field H_d ($= N_d M$) which is, in fact, proportional to M and directed opposite to it. Here N_d is the linear demagnetizing factor. The sum of the three orthogonal vectors is equal to one ($N_x + N_y + N_z = 1$). The values for the three vectors depend on the shape of the magnetic unit. An important factor is that the magnitude of the internal field (H_{int}) is less than the applied field (H_{app}) by the amount of the H_d . Consequently the H_{app} must overcome the H_d in order to saturate the material.

Magnetization of Deposited Alloys. Frequently the relation between the magnetization and composition is presented by the Slater-Pauling curve which gives the relation between the saturation magnetization (M_s) and homogeneous bulk Co–X (78,79). In general, most papers report that the M_s of sputtered and evaporated films, deposited at higher substrate temperatures, is found to be larger than that for bulk alloys having the same average chemical composition. Although in the literature various origins have been proposed, the most likely explanation is the phase separation. Two hcp phases, which hardly occur in bulk Co–Cr material at low temperatures, cause compositional fluctuations along the grain boundaries as the film growth proceeds.

Figure 19 compares experimental data with calculated curves (80). In the random Co–Cr alloy the Cr atoms are not distributed in the most suitable way for reducing the M_s of the alloy. Therefore the maximum local content of Cr for this distribution is much higher than in the case where Cr–Cr bonds are not present. The curve for no Cr–Cr bonds present shows that the M_s becomes zero at 25 at. % Cr, based on the fact that for bulk material the measured M_s for this composition is zero. Consequently 4 Cr nearest neighbor in an hcp lattice makes the final M_s zero.

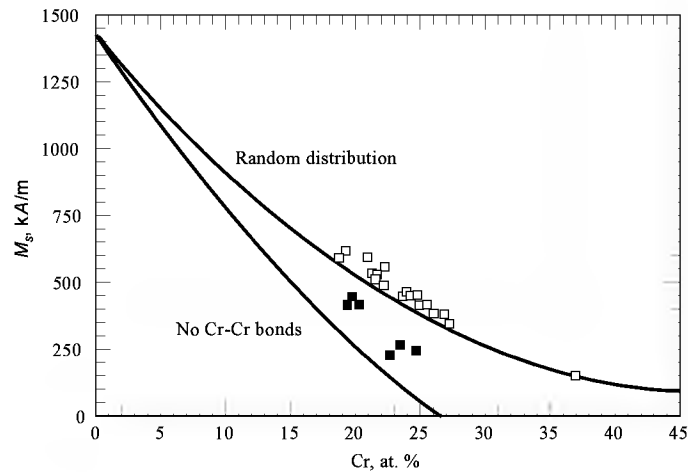


Fig. 19. M_s as a function of the Cr content in Co: (□), coevaporation at low temperature; (■), sputtered; and (—), calculated (80).

Coercivity of Thin-Film Media. The coercivity in a magnetic material is an important parameter for applications but it is difficult to understand its physical background. It can be varied from nearly zero to more than 2000 kA m in a variety of materials. For thin-film recording media, values of more than 250 kA m have been reported. First of all the coercivity is an extrinsic parameter and is strongly influenced by the microstructural properties of the layer such as crystal size and shape, composition, and texture. These properties are directly related to the preparation conditions. Material choice and chemical inhomogeneties are responsible for the M_s of a material and this is also an influencing parameter of the final H_c . In crystalline material, the crystalline anisotropy field plays an important role. It is difficult to discriminate between all these parameters and to understand the coercivity origin in the different thin-film materials in detail.

As has been seen in an ideal single-domain particle which switches coherently, the H_c is equal to the H_k if a field is applied in the easy axis direction. Depending on various factors, incoherent switching occurs and the H_c decreases. Even in a matrix of particles, with magnetostatic interaction the coercivity will again be influenced. Multidomain particles and thin films switch by domain-wall motion and again the coercivity decreases in comparison with the H_k . For many different morphologies the H_c is quite well described by the following general relation (81): $H_c = \alpha(2K_1 / M_s) - N_{eff}^* M_s$ where N_{eff} = effective demagnetization factor. This relation gives a modification of the magnetocrystalline anisotropy and demagnetizing fields. The parameters α and N_{eff} relate to the microstructural aspects and the dimensions of the magnetic units. The influences of α and N_{eff} on the H_c for polycrystalline thin films have been discussed (82). In addition, for thin films surface and interface properties also influence the H_c . This leads to the assumption that the coercivity (an extrinsic property) can only be determined by means of the macroscopical hysteresis loop in combination with the theory of micromagnetism (83). Knowledge of the microstructural properties of the material cannot be omitted. In the first place the size and shape of the magnetic unit (crystal, column) material plays an important role. In single-domain particles the coercivity increases to a maximum at a critical particle-size diameter. Future increase of the diameter results in a multidomain-state particle. Decreasing the single-domain particle diameter it finally becomes superparamagnetic. At this stage H_c becomes independent of the diameter and the reversal takes place by thermal activation which again leads to a lower H_c . In the case of a multidomain particle the H_c is determined by pinning mechanisms of the domain wall. These mechanisms are determined by the magnetically inhomogeneous regions like columnar boundaries, chemical inhomogeneties, stacking faults, etc.

Thin-Film Media for Various Types of Recording

Thin-film media can be made by various technologies, eg, sputtered deposited Co–Cr–X films for longitudinal applications, laminated media for hard disk application, metal evaporated tape, and multilayers for possible applications in magneto optic recording.

Sputtered Co–Cr–X/Cr Disks for Longitudinal Recording. Sputtered hard disks of Co–Cr–X ($X = \text{Ni, Ta, Pt, C}$) for very high density longitudinal recording have been prepared and tested by many industrial and university groups. The first disks with a density larger than 1 Gbit/in.² have been reported by IBM (13) and Hitachi (Japan) (16). The recording layers are deposited on a Cr underlayer which is deposited on the hard disk substrate. The composition is a Co ternary or quaternary composition and sputtering is the most favorable deposition technology for preparing these disks. The substrate consists of Ni:P plated on Al, glass, or canasite. The most important process parameters for depositing a suitable magnetic layer are the composition, thickness, and structure of the layer. The latter also depends on the structure of the Cr underlayer. The deposition parameters for both layers such as substrate temperature, bias voltage, and surface texture are also important.

Choice of Co–Cr. Sputtered Co–Cr films for perpendicular magnetic recording were first made at the Tohoku University in 1975 (84,85). An overview of the preparation, microstructure, and magnetic properties of Co–Cr thin films has been given (49). Before Co–Cr was chosen, many different alloys of Co– M materials were investigated (86,87), including $X = \text{Rh, Pd, Mo, W, V, Ti, Cr, Mo, Pt, and Mn}$. All the Co– M films exhibit an hcp crystal structure with the c -axis oriented perpendicular to the film surface deposited on a variety of substrate materials. However if a Cr layer is used as seed layer (or buffer layer) then the Co–Cr start growing with an in-plane c -axis which is a condition for longitudinal recording. For further tailoring longitudinal media, ternary and quaternary alloys have been prepared and used as high density media (14,16).

Role of the Cr Underlayer. One of the methods to overcome the problem of low coercivity in vacuum evaporated films is an underlayer between the substrate and the ferromagnetic layer, which was already proposed in 1967 (88). By slow deposition (0.1 nm/s) of a Co layer (thickness less than 100 nm) on a Cr underlayer with bcc structure the coercivity increases at a level suitable for magnetic recording. The H_c is strongly dependent on the rate of deposition, the thickness of the Co as well as the Cr layers, and the substrate temperature. The Cr underlayer induces the growth of the Co layer with an exclusive hexagonal crystalline structure and a narrow crystallite-size distribution. If the Cr layer increases, its crystal size also increases and the Co grows more quasisepitaxial. The c -axis orientation of the Co becomes more in-plane if the substrate temperature increases. An important fact is that this effect is strongest if the deposition of the Co occurs immediately after the deposition of the Cr underlayer (no oxidation). The role of the Cr underlayer is thus the creation of the right conditions for epitaxial growth of the polycrystalline layer having the hcp texture c -axis in the plane of the medium. Optimizing the Cr layer also controls the crystal size and morphology. It was reported in 1986 (89,90) that the Cr underlayer thickness has a great influence on the coercivity of the Co–Ni–Cr layer. In most of the literature it can be found that with increasing Cr thickness the H_c increases. Under ideal conditions and the right material combinations coercivities above 240 kA/m have been prepared.

Randomly In-Plane Texture. The c -axis orientation is important for the magnetic anisotropy. The Co-based alloys used mostly have an hcp structure with the c -axis randomly in-plane, but it is also possible to create a tilted c -axis depending on the underlayer structure (91). On thick Cr underlayers the (10.1) Co–Cr–Ta orientation growth on the Cr (100) plane finally leads to a tilted c -axis of 28° with respect to the substrate plane, but, for instance, the (11.0) of Co–Cr–Pt layers grow epitaxially on the Cr (100) resulting in a c -axis orientation of the magnetic layer in-plane. An optimized Cr thickness is found for this type of epitaxial growth. Much research will have to be carried out before a complete understanding of this type of growth can be obtained.

Oriented In-Plane Texture. In this kind of film the properties (H_c and M_r) in the various in-plane directions (texture and nontexture directions) are different. The texture of the film can be supported by the texture of the substrate and the crystal lattice can be smaller in the texture direction than in the transverse direction. This can be the source for strain-induced magnetic anisotropy (magnetostriction). It is also found that the crystal is aligned in the texture direction (92).

Morphology. Thin films should magnetically operate as individual particles. The exchange between the crystals should be broken or at least minimized. This can be realized by mechanical separation or by segregated crystal boundaries with nonferromagnetic compositions. The Cr underlayer is responsible for the physical separation. Cr grows with a columnar morphology and has a rough surface structure (93). It was reported that the dome-like surface roughness is on the order of tens of nm depending on the thickness of the Cr layer. It can be observed from TEM observations (13) that in combination with a high Ar sputter-gas pressure the Co–Cr–Pt layer shows physically separated crystal boundaries.

Compositional Separation. Similar effects as shown for Co–Cr films used for the perpendicular recording mode compositional separation also occurs in Co–Cr–X films for longitudinal recording, although at present only limited experimental data is available. Most data from in-plane media is derived from macroscopic magnetic measurements, which on Co–Cr–Ta showed that an increase in Ta content also increases the magnetization and suggests that this effect is due to increasing the Cr segregation at the boundaries. Another effect by applying Ta is the increase of the lattice constant which leads or reduces strain in the lattice, and it was suggested that this influences the H_c . In Co–Cr–Pt it was found that the increase of the lattice constant can be correlated with the increase of H_c . The Pt in these types of ternary alloys can also play a role by the formation of the very hard Co₅₀Pt₅₀ composites which have a large influence on the reversal behavior.

Laminated Hard Disk Media. Many studies have been published on thin-film media consisting of more magnetic layers separated by a nonferromagnetic interlayer (94–98). The main aim for studying the laminated medium structures for high density recording is to improve the signal-to-noise ratio by reducing the medium noise. The linear density is determined by the macroscopic magnetic properties H_c and remanence thickness product $M_r\delta$ while the medium noise is related to the individual grain structure and the exchange coupling between the grains.

Individual control of the parameters H_c and the $M_r\delta$ in thin films was announced in 1979 (99) proposing stacked Co-films. The increase of H_c and $M_r\delta$ in a four-layer system consisting of a Cr underlayer (300 nm) CoNiCr (35 nm) Cr interlayer (12.5 nm) CoNiCr (35 nm) have shown better read/write characteristics (100). It has also been reported that the noise contribution in such a system can be reduced because the noise of the individual layers is lower than that of a thicker single layer (96). The total noise is the sum of the noise of the individual two layers plus a cross term. The key is now to design individual layers with low noise behavior. One of the advantages of the laminated technology is that the growth of the magnetic layer is interrupted in an early stage which gives a better opportunity to obtain uncoupled (exchange) crystals. This research is in progress by using a variety of materials and deposition parameters. An important aspect is that the interlayer thickness should be thick enough for interrupting the exchange between the magnetic layers but thin enough to keep magnetostatic coupling. In comparison with the common technologies the manufacturing control of macromagnetic and micromagnetic properties is more difficult in the laminated process (101).

Metal Evaporated Tape. Very pure films and to a certain extent preselected structures and morphologies can be obtained by vacuum evaporation. Atoms and molecules are emitted from the sources by heating and exist in a gaseous state. The pressure in the vacuum chamber, a certain equilibrium pressure (saturated-vapor pressure), is established at a given temperature. The deposition rate (R) depends on the vapor pressure. The best results can be obtained if the evaporated elements and their alloys have a similar vapor pressure, but this is a limitation of the method. In the case of deposition of Co–Ni as a recording media this problem is not present. There are two methods to overcome the low H_c by evaporation, namely a Cr underlayer and deposition by varying the angle of incidence of the arriving metal atoms (102).

In order to produce the so-called ME (metal evaporated) tape, modifications relative to the oblique incidence evaporation have been made in order to obtain efficient use of the evaporated material and continuous evaporation over the total length of the substrate, ie, hundreds of meters of tape have to be produced in one single run at a very high production speed (Fig. 20). The most important modification is that the evaporation occurs over a range of angles, instead of one angle, whereby the incidence angle changes as the substrate passes along the vapor beam. This means that the direction of the elongated columns changes throughout the thickness of the tape. The properties of continuous varied incidence Co, Ni, Co–Ni, and Fe tapes have been investigated (103).

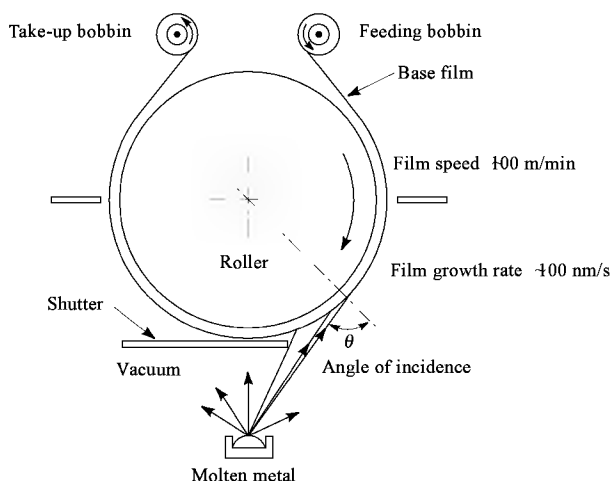


Fig. 20. Mass-production equipment for metal evaporated tape (103); vacuum ~ 1 mPa (10^{-3} torr).

A principle deposition geometry is given in Figure 20. Depending on the rotation direction of the substrate (tape) two nucleation processes can take place, namely HIN and LIN. In high incidence-angle nucleation (HIN) the nucleation takes place at high incoming angles of the beam and successive growth occurs with a decreasing incidence angle. In low incidence-angle nucleation (LIN) the nucleation starts at low incidence angles and further growth occurs with an increase of the incidence angle. But at such a high angle of incidence the incoming vapor beam is almost parallel to the substrate and the deposition rate becomes very low. In addition the packing fraction is strongly reduced.

Last but not least the corrosion resistance of this type of material is very poor. A partly oxidized medium is the solution; therefore during deposition oxygen was added. The final composition of Co-Ni-O also gave the right magnetic medium properties. The H_c is strongly influenced by the O_2 value.

The morphology of the ME tape can be seen in Figure 21. In the cross-sectional microstructure, prepared by ion milling, the so-called banana-shaped morphology is clearly visible. Due to the shadowing effect the density of such a film is not high and the micromagnetic behavior is probably a mixture of domain-wall motion and rotation of the magnetization. The structural analysis of the tape provides the following data. The thickness of the magnetic layer is about 130 nm and a regular structure composed of very fine fibers is observed (thickness 3–10 nm). The columns make an angle of 37° relative to the plane of the film. Auger electron spectroscopy (aes) has shown an average composition of $Co_{77}Ni_{10}O_{13}$ (104,105).

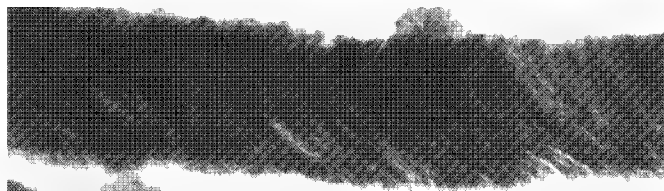


Fig. 21. Transmission electron micrograph (tem) made from a cross-section of a metal-evaporated tape medium with the typical banana-like shape of columns. The layer thickness is ~ 120 nm.

The substrate side seems to be less dense which is to be expected from large shadowing at a high incidence angle. It is anticipated that the fiber-like structure consists of crystallites. An hcp Co phase as well as a Ni-O fcc phase were found by x-ray diffraction. The anisotropy of such a complicated composition has been analyzed (106) and a correlation between the anisotropy direction and pulse shape have also been analyzed (105). The angle of incidence method has also been used for a hard disk preparation (107). Two electron gun sources were used and the evaporated beam was deposited under an angle of 60° on both sides of the rotatable-disk substrate. The material used was $Fe_{42.5}Co_{42.5}Cr_{15}$ and was deposited on a substrate of Al-Mg with an underlayer of stainless steel which needs a special treatment to create the right surface structure of the magnetic layer to prevent adhesive contact between the head and the medium.

Multilayers for Magneto-optic Recording. In 1957 the first MO medium, a thin film of MnBi, was proposed and based on the principle of Curie temperature writing. Later MnAlGe and CoPt were intensively studied followed by GdCo, but this material was used for compensation-point writing. None of the materials was ideal; for instance MnBi consists of two crystallographic phases and is very unstable above 400°C . The noise problem was high in the case of MnAlGe media and domain instability played the most important role in the case of GdCo.

Magneto-optic materials have been selected on the basis that they can sustain submicrometer domains, have a large signal-to-noise ratio supported by sufficient magneto-optic effects, and are resistant to corrosion and have chemical stability. The most popular materials at the present time are the amorphous rare-earth (Gd, Tb, and Dy) transition-metal (Fe, Co) alloys which are deposited by evaporation or sputtering alloys like GdTbFe and TbFeCo are well known and can be made commercially. There is a great interest in and a great deal of research on so-called multilayer structures (see Fig. 16) for MO applications (108).

Mainly Co-Pt and Co-Pd have been studied by evaporation and sputtering. Although both processes show different physical processes the MO properties of the films do not vary much. Studies have also been carried out for materials with a lower Curie temperature (109).

Conclusions

Particulate recording media have been prepared with success since the 1950s although they have shown disadvantages. Thin-film hard disk materials are attracting great interest for ultrahigh density recording either for longitudinal or perpendicular modes. There is a trend to use more and more thin-film heads; also, the MR head is used for many high density recording set-ups. Ternary and quaternary Co-Cr-X-Y alloys have been proposed as suitable media for longitudinal high density thin-film media. Underlayers (mostly Cr) are necessary for controlling the magnetic properties. Higher density in longitudinal recording needs thinner layers and smaller crystallites. A trend is that the layer thickness for hard disk application must decrease to about 10–15 nm for obtaining 10 Gbit/in.² density recording. Thin-film materials are also essential for magneto-optic applications, and an increase in densities can also be expected from the various multilayers developed so far.

The microstructure and morphology is important for the new class of high density recording media. Magnetic properties which should be optimized are the coercivity, the remanence-thickness product ($M_r\delta$), and the S^* . The sharpness and shape of the written transition is determined by the grain morphology, magnetical decoupling of the grains, and their textures. The higher M_r and lower δ for obtaining ultimate density will be finally limited by the superparamagnetism. High coercivity and low noise media are required for longitudinal recording. In the case of the perpendicular mode of recording optimization has to be carried out for the normalized-particle coercivity distribution as proposed in Reference 110.

A trend from the history of the high density recording technology shows that the recording density has increased by a factor 10^3 since the 1960s. Densities of 10 Gbit in.² for hard-disk applications and several bits μm for tape application, are part of long-term planning. With respect to perpendicular recording, high linear-bit densities have been demonstrated on laboratory scale, achieved by recording a medium consisting of Co–Cr–Ta on a soft magnetic underlayer with a special single-pole head (111). In the case of pmr the ultimate densities of bit areas (2500 nm²) are predicted with bit lengths smaller than 50 nm (10).

BIBLIOGRAPHY

1. C. D. Mee and E. D. Daniel, *Magnetic Recording*, Vols. 1–3, McGraw-Hill Book Co., Inc., 1987–1988; and updated ed., *Magnetic Recording Handbook, Technology and Application*, 1990.
2. A. S. Hoagland and J. E. Monson, *Digital Magnetic Recording*, 2nd ed., John Wiley & Sons, Inc., New York, 1991.
3. R. M. White, ed., *Introduction to Magnetic Recording*, IEEE Press, New York, 1985.
4. J. C. Mallinson, *The Foundations of Magnetic Recording*, 2nd ed., Academic Press, Inc., London, 1993.
5. K. H. J. Buschow and co-workers, eds., *High Density Digital Recording*, Kluwer Academic Publishers, Series E: Applied Sciences-Vol. 229, NATO ASI series, Dordrecht, the Netherlands, 1993.
6. T. C. Arnoldussen and co-workers, eds., *Noise in Digital Magnetic Recording*, World Scientific Publishing Co., Singapore, 1992.
7. H. N. Bertram in J.-G. Zhu, in H. Ehrenreich and D. Turnbull, eds., *Solid State Physics*, 46, Academic Press, Inc., San Diego, 1992, pp. 271–371.
8. B. Bhushan, *Tribology and Mechanics of Magnetic Storage Devices*, Springer-Verlag, New York, 1990.
9. H. N. Bertram, *Theory of Magnetic Recording*, Cambridge University Press, Cambridge, U.K., 1994.
10. Y. Nakamura, *J. Magn. Soc. Jpn.* **15**(52), suppl., (1991).
11. I. Tagawa and Y. Nakamura, *J. Magn. Soc. Jpn.* **13**(S1), 97 (1989).
12. E. S. Murdock, R. F. Simmons, and R. Davidson, *IEEE Trans. Magn.* **28**, 3078 (1992).
13. T. Yogi and co-workers, *IEEE Trans. Magn.* **26**(5), 2271 (1990).
14. C. Tsang and co-workers, *IEEE Trans. Magn.* **26**, 1689 (1990).
15. T. D. Howell and co-workers, *IEEE Trans. Magn.* **26**, 2298 (1990).
16. M. Futamoto and co-workers, *IEEE Trans. Magn.* **27**, 5280 (1991).
17. M. L. Williams and R. L. Comstock, *AIP Conf. Procs.* **5**, 378 (1971).
18. D. B. Richards and T. J. Szczech, *J. Appl. Phys.* **49**(3), 1819 (1978).
19. E. Koster, H. Jakusch, and U. Kullmann, *IEEE Trans. Magn.* **17**(6), 2250 (1981).
20. T. Suzuki, *IEEE Trans. Magn.* **20**(5), 675 (1984).
21. S. Iwasaki and K. Tekamura, *IEEE Trans. Magn.* **11**, 1173 (1975).
22. E. Katz and P. Schreiber, *J. Magn. Soc. Jpn.* **15**(S1), suppl., 483 (1989).
23. H. Hamilton, *J. Magn. Soc. Jpn.* **15**(S2), suppl., 483 (1991).
24. Proceedings of the PMRC '89, *J. Magn. Soc. Jpn* **13**(S1), suppl., more than 900 pages (1989), and Proceedings PMRC '91, *J. Magn. Soc. Jpn.* **15**(S2), suppl., more than 1000 pages (1991).
25. S. Iwasaki, Y. Nakamura, and K. Ouchi, *IEEE Trans. Magn.* **15**, 1456 (1979).
26. T. Suzuki, *IEEE Trans. Magn.* **20**, 675 (1984).
27. C. Kooy and U. Enz, *Philips Res. Repts.* **15**, 7 (1960).
28. H. J. Williams and co-workers, *J. Appl. Phys.* **28**, 1181 (1957).
29. F. J. A. M. Greidanus, *Philips J. of Res.* **45**, 19 (1990).
30. W. Bas Zeper, *Magneto-optical Recording Media on Co/Pt Multilayers*, Ph.D. dissertation, University of Twente, Enschede, the Netherlands, 1991.
31. W. Reim, J. Schoenes, and P. Wachter, *IEEE Trans. Magn.* **20**, 1045 (1984).
32. A. Aharoni, *IEEE Trans. Magn.* **22**, 478 (1986).
33. E. Stoner and E. P. Wohlfarth, *Philos. Trans. R. Soc. Lond. Ser. A* **240**, 74 (1948).
34. I. S. Jacobs and C. P. Bean, *Phys. Rev.* **100**, 1060 (1955).
35. C. P. Bean, *J. Appl. Phys.* **26**(11), 1381 (1955).
36. S. D. Cullity, *Introduction to Magnetic Materials*, Addison-Wesley, Reading, Mass., 1972, pp. 399–418.
37. T. Wielinga, J. C. Lodder, and J. Worst, *IEEE Trans. Magn.* **18**, 1107 (1982).
38. R. F. Bunshan and co-workers, *Deposition Technologies for Films and Coatings*, Noyes Data Corp., Park Ridge, N.J., 1982.
39. L. I. Maisel and R. Clang, eds., *Handbook of Thin-Film Technology*, McGraw-Hill Book Co., Inc., New York, 1970.
40. D. O. Smith, M. S. Cohen, and G. P. Weiss, *J. Appl. Phys.* **31**, 1755 (1960).
41. J. G. W. van de Waterbeemd and G. W. van Oosterhout, *Philips Research Reports* **22**, 375 (1967).
42. A. G. Dirks and H. J. Leamy, *Thin Solid Films* **47**, 219 (1977).
43. A. G. Dirks and H. J. Leamy, *J. Appl. Phys.* **49**(6), 2430 (1978).
44. H. Fujiwara and co-workers, *Thin Solid Films* **163**, 387 (1988).
45. J. M. Nieuwenhuizen and H. B. Haanstra, *Philips Tech. Rev.* **27**(1), 87 (1966).
46. S. Keitoku, *J. Sci. Hiroshima Univ., Ser. A* **37**(2), 167 (1973).
47. H. van Kranenburg and J. C. Lodder, *Mat. Sci. Eng. Rep.* **R11**(7), 295–354 (1994).
48. B. Chapman, *Glow Discharge Processes*, John Wiley & Sons, Inc., New York, 1980.
49. J. C. Lodder, in K. H. J. Buschow and co-workers, eds., *High Density Recording*, NATO ASI Applied Sciences-vol. 229, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1993, Chapt. 6, pp. 59–99.
50. B. A. Movchan and A. V. Demshishin, *Phys. Met. Metallogr.* **28**, 83 (1969).
51. J. A. Thornton, *Am. Rev. Mater. Sci.* **7**, 239 (1977).
52. R. Messier and co-workers, *J. Vac. Sci. Techn.* **A2**(2), 500 (1984).
53. M. Aspland, G. A. Jones, and B. K. Middleton, *IEEE Trans. Magn.* **5**, 314 (1969).
54. J. Smits, S. B. Luitjens, and F. J. A. den Broeder, *J. Appl. Phys.* **55**(6), 2260 (1984).
55. A. K. Jinghan, *J. Magn. Magn. Mater.* **54–57**, 1685 (1986).
56. U. Hwang and co-workers, *Thin Solid Films* **147**, 231 (1987).
57. J. N. Chapman, I. R. McFadden, and J. P. C. Bernards, *J. Magn. Magn. Mater.* **62**, 358 (1986).
58. M. Sagoi, R. Nishikawa, and T. Suzuki, *IEEE Trans. Magn.* **22**(5), 1335 (1986).
59. F. F. Abraham and C. R. Bundle, *J. Vac. Sci. Techn.* **18**(2), 506 (1981).
60. K. Ishida and T. Nishizawa, *Bull. Alloy Phase Diag.* **11**, 357 (1990).
61. N. L. Peterson, *J. Vac. Sci. Techn. A* **4**(6), 3066 (1986).
62. Y. Maeda, S. Hirono, and M. Asahi, *Jpn. J. Appl. Phys.* **24**, L951 (1985).
63. Y. Maeda and co-workers, *J. Magn. Soc. Jpn.* **15**(S2), suppl., 457 (1989).
64. Y. Maeda and M. Asahi, *J. Appl. Phys.* **61**(5), 1972 (1987).
65. K. Hono and co-workers, *Appl. Phys. Lett.* **62**, 2504 (1993).

66. Y. Maeda and co-workers, *J. Magn. Soc. Jpn.* **15**(S2), 457 (1991).
67. Y. Maeda and K. Takei, *J. Magn. Soc. Jpn.* **16**(2) (1992).
68. K. Ishida and T., Nishizawa, *Bull. Alloy Phase Diag.* **11**, 357 (1990).
69. T. Wielinga, *Investigations on Perpendicular Recording*, Ph.D. dissertation, University of Twente, Enschede, the Netherlands, 1983.
70. J. A. Aboaf and E. Klokholm, *IEEE Trans. Magn.* **17**(6), 3160 (1981).
71. K. Kobayashi and G. Ishida, *J. Appl. Phys.* **52**(3), 2453 (1981).
72. T. Yamashita and co-workers, *IEEE Trans. Magn.* **27**, 4727 (1991).
73. C. R. Paik and co-workers, *IEEE Trans. Magn.* **28**(5), 3084 (1992).
74. O. Lopez and D. A. Clark, *J. Appl. Phys.* **57**(1), 3943 (1985).
75. O. Lopez and D. A. Clark, *IEEE Trans. Magn.* **21**(5), 1401 (1985).
76. Y. Nakamura, *J. Magn. Soc. Jpn.* **13**(S1), 33 (1989).
77. W. J. M. A. Geerts, J. C. Lodder, and Th. J. A. Popma, *J. Magn. Magn. Mat.* **104–107**, 971 (1992).
78. R. M. Bozorth, *Ferromagnetism*, Van Nostrand, New York, 1951, p. 441.
79. J. C. Slater, *J. Appl. Phys.* **8**, 385 (1937), and L. Pauling, *Phys. Rev.* **54**, 899 (1938).
80. W. G. Haines, *J. Appl. Phys.* **55**(6), 2263 (1984).
81. H. Kronmüller, *Phys. Stat. Sol.(b)* **144**, 385 (1987).
82. H. Kronmüller, *J. Magn. Soc. Jpn.* **17**, 260 (1993).
83. W. F. Brown, *Micromagnetics*, Wiley-Interscience, New York, 1963, p. 75.
84. S. Iwasaki and H. Yamazaki, *Proc. of the 7th Conf. on Magn. Soc. Japan*, 4pA-7, (1975), in Japanese.
85. S. Iwasaki and K. Ouchi, *IEEE Trans. Magn.* **14**, 849 (1987).
86. S. Iwasaki, K. Ouchi, and N. Honda, *IEEE Trans. Magn.* **16**(5) (1980).
87. K. Kobayashi and G. Ishida, *J. Appl. Phys.* **52**(3), 2453 (1981).
88. J. P. Lazzari, I. Melnick, and D. Randet, *IEEE Trans. on Magn.* **3**(3), 205 (1967).
89. G. Chen, *IEEE Trans. Magn.* **22**, 334 (1986).
90. M. Ischikawa, *IEEE Trans. Magn.* **22**(5), 573–575 (1986).
91. G. E. Johnson and co-workers, *J. Appl. Phys.* **67**, 4686 (1990).
92. J. M. Simpson and co-workers, *IEEE Trans. Mag.* **23**, 3405 (1987).
93. S. Agarwal, *IEEE Trans. Magn.* **21**, 1527 (1986).
94. S. E. Lambert and co-workers, *IEEE Trans. Magn.* **29**(1), 223–229 (1993).
95. S. E. Lambert, J. K. Howard, and I. L. Sander, *IEEE Trans. Magn.* **26**(5), 2706–2708 (1990).
96. E. S. Murdock, B. R. Natarajan, and R. G. Walmsley, *IEEE Trans. Magn.* **26**(5), 2700 (1990).
97. H. T. Hata and co-workers, *J. Appl. Phys.* **67**, 642 (1990).
98. D. C. Palmer and co-workers, *IEEE Trans. Magn.* **27**, 307 (1991).
99. W. T. Maloney, *IEEE Trans. Magn.* **15**(3), 1135 (1979).
100. S. Katayama and co-workers, *IEEE Trans. Magn.* **24**(6), 2982–2984 (1988).
101. K. E. Johnson and co-workers, *IEEE Trans. Magn.* **29**(1), 215 (1993).
102. D. E. Speliotis and co-workers, *J. Appl. Phys.* **36**, 972 (1965).
103. K. Nakamura, Y. Ohta, and A. Itho, *IEEE Trans. Magn. Mag.* **18**, 1077 (1982).
104. J. S. Gau and co-workers, *J. Appl. Phys.* **61**(8), 3807 (1987).
105. G. Krijnen and co-workers, *IEEE Trans. Magn.* **24**(2) (1988).
106. S. Swaving and co-workers, *J. Magn. Magn. Mater.* **67**, 155 (1987).
107. T. C. Arnoldussen and co-workers, *IEEE Trans. on Magn. Mag.* **20**, 821 (1984).
108. C. J. Robinson, T. Suzuki, and C. M. Falco, eds., *Materials for Magnet-Optic Data Storage*, Materials Research Society Proceedings 150, Pittsburgh, Pa., 1989.
109. M. Mes and co-workers, *J. Magn. Soc. Jpn.* **17**, 44–47 (1993).
110. K. Tagawa and Y. Nakamura, *IEEE Trans. Magn.* **27**(6), 4975 (1991).
111. Y. Nakamura and H. Muraoka, *IEEE Trans. Magn.* **27**, 4555 (1991).

J. C. Lodder
University of Twente

INFRARED DETECTORS.

See INFRARED TECHNOLOGY AND RAMAN SPECTROSCOPY; PHOTODETECTORS.

INFRARED TECHNOLOGY AND RAMAN SPECTROSCOPY

Infrared technology
Raman spectroscopy

INFRARED TECHNOLOGY

The infrared (ir) region of the electromagnetic spectrum has been used for a great variety of purposes, from radiant heating to fiber optic communications (see FIBER OPTICS). The applications most relevant to chemical technology are spectrometry and radiometry. Spectrometry is a molecular analysis method based on measuring the wavelength dependent interaction of radiation with matter. Radiometry measures the amount of radiation, thereby remotely monitoring thermal factors. The ir region of the electromagnetic spectrum lies between the visible and microwave regions, and is generally taken to be from 0.75 to 1000 μm (13,333 to 10 cm^{-1}). This wide expanse is further divided by spectroscopists into the near infrared, 0.75 to 2.5 μm (13,333 to 4,000 cm^{-1}); the mid-infrared, 2.5 to 25 μm (4000 to 400 cm^{-1}); and the far infrared, 25 to 1000 μm (400 to 10 cm^{-1}). These divisions roughly correspond to regions having different kinds of spectral features and involving different kinds of equipment. For radiometry the ir spectrum is divided up differently. The short-wave infrared (swir) extends from 1 to 3 μm , the mid-wave infrared (mwir) from 3 to 8 μm , and the long-wave infrared (lwir) from 8 to 13 μm .

Equipment

Sources. Although broad-band sources are based on thermal emission, few produce a true blackbody emission spectrum. Blackbody sources are used for the calibration of radiometric equipment and as standards for determining absolute emissivities or reflectivities. A small-aperture blackbody source consists of a large cavity at uniform temperature having a small hole in one wall. Commercial small-aperture sources have apertures under 25 mm in diameter, operating temperatures between 6 K and 3000°C, and emissivities as high as 0.999 ± 0.0005 . Large-aperture sources, essentially enclosures in which part of one wall is missing, are used for calibrating thermal imagers. The open nature of such sources reduces precision and raises power requirements relative to small-aperture sources. Available sources have apertures as large as 4000 cm^2 , operating temperatures up to 600°C, and emissivities exceeding 0.99.

The optimum source for spectrometry depends on the spectral region. A quartz halogen lamp having a tungsten filament is excellent for the near infrared region. Its filament is a 3000 K emission source having an emissivity near 0.7. It is only useful to about 3 μm , at which point the quartz envelope becomes opaque. Because there are no suitable envelope materials for longer wavelengths, mid-infrared sources must function exposed to air, which reduces attainable source temperatures. The three common types of mid-infrared sources are globars, Nernst glowers, and nichrome coils. Globars are silicon carbide (carborundum), [409-21-2] SiC, helices or rods, typically about 6 mm wide, resistively heated up to 1100°C. Globars are robust, inexpensive, and radiate with an emissivity of about 0.8 out to 100 cm^{-1} , so they can be used in the far infrared region as well. These typically use 100 to 200 W and are water cooled. Nernst glowers resistively heat rare earth oxides, principally zirconia, to 1100 to 1500°C. Nernst glowers are narrow (1 to 3 mm) air-cooled sources that use 20 to 40 W. Although Nernst glowers run hotter than globars, the lower emissivity of the glowers at short wavelengths greatly reduces the irradiance difference. Nichrome wire coils are the simplest and cheapest of mid-infrared sources. These operate in the 1000 to 1100°C range, are air cooled, and are somewhat less stable than globars and Nernst glowers. The mid-infrared sources can be used for the high energy end of the far infrared region, but any constant-emissivity thermal source hot enough to be useful at far infrared wavelengths has the majority of its output in the mid-infrared region. The high pressure mercury arc in a quartz envelope is the preferred source for wavelengths beyond 100 μm . The quartz envelope is opaque between 3 and 100 μm and acts as a 750°C radiator over that range. At longer wavelengths the quartz is transparent and the internal plasma, which acts as a 10,000°C blackbody, is the radiation source (1).

Lasers (qv) are the principal monochromatic sources in the infrared region. The most ubiquitous is the diode laser. Most diode lasers are based on the edge-emitting double heterojunction (or heterostructure). In such a laser, a thin active layer of compound semiconductor is sandwiched between two different semiconductor materials, one p-type, the other n-type. The ends of the sandwich are polished to act as partial reflectors and form the laser cavity. The typical active layer is 1 μm thick, 1 to 10 μm wide, and 500 μm long. When the diode is forward biased, electrons from the n-type material and holes from the p-type material enter the active layer, combine, and emit photons, so the bandgap of the active layer governs the wavelength emitted (see also LIGHT GENERATION, SEMICONDUCTING LASERS).

Most commercial diode lasers are based on the Group 13–15 (III–V) compound gallium arsenside, [1303-00-0] GaAs, in which some of the gallium, arsenic, or both have been replaced by other Group 13 (III) or Group 15 (V) elements to produce the desired bandgap. The material $\text{In}_{1-x}\text{Ga}_x\text{As}_y\text{P}_{1-y}$ can lase anywhere within the range 1.06 to 1.65 μm , depending on x and y . The silica fiber optics used by the telecommunications industry have their lowest loss near 1.55 μm and their zero group-velocity dispersion near 1.31 μm , so most InGaAsP diode lasers are either $\text{In}_{0.73}\text{Ga}_{0.27}\text{As}_{0.58}\text{P}_{0.42}$ for 1.31 μm or $\text{In}_{0.58}\text{Ga}_{0.42}\text{As}_{0.9}\text{P}_{0.1}$ for 1.55 μm . Other infrared laser diode materials include GaAlAs for the 0.62 to 0.905 μm range, PbEuSeTe for 3.3 to 5.8 μm , PbSSe for 4.2 to 8.0 μm , PbSnTe for 6.3 to 29 μm , and PbSnSe for 8.0 to 29 μm . Substituting one element for another within the same chemical group changes the bandgap by changing the lattice constant of the crystal. The larger the constant, the smaller the bandgap. Some diode lasers can be tuned by changing the temperature or pressure, because these also affect the lattice constant. The tuning, however, is discontinuous; as the temperature (or pressure) rises, a laser tunes continuously over a range on the order of 0.1% of the nominal wavelength, then it mode-hops to a new wavelength from which it continuously tunes until it again hops. Infrared light-emitting diodes are made from the same materials as lasers (2,3) (see LIGHT GENERATION, LIGHT EMITTING DIODES).

The exact wavelength of the neodymium-ion laser depends on the matrix in which the Nd^{3+} is placed. For example, Nd:YAG (yttrium aluminum garnet) lasers function at 1.064 μm and are good for high repetition rate applications because of the good thermal characteristics of YAG. Versions operating continuously or up to 50 kHz pulsed are available having average outputs up to 1.8 kW. Nd:glass lasers are more efficient but have poorer thermal characteristics. These are used for high (up to 400 J) pulse energy, low pulse rate applications. Numerous other fixed wavelength lasers based on an ion-doped solid matrix are available. Vibronic solid-state lasers are very similar in that these operate by optically pumping an ion doped in a crystal, but are

tuneable because of numerous overlapping vibrational levels. These include Ti:sapphire with a range of 0.67 to 1.13 μm, alexandrite (Cr:BeAl₂O₄) for 0.72 to 0.80 μm, and Co:MgF₂ for 1.75 to 2.5 μm. Color-center (or F-center) lasers are similar to these vibronic lasers, but the optically active species is a crystal defect. These consist of doped and x-ray treated alkali-halide crystals that must be kept below 200 K. Commercial color-center lasers are tuneable over the 1.45 to 1.75 μm and 2.3 to 3.45 μm ranges (4).

Dye lasers produce laser action from organic dyes that are dissolved in a liquid solvent and driven by a flashlamp or another laser. These are principally used in the visible spectrum, but dyes are available out to 1.32 μm. The tuning range of a single dye is 0.10 to 0.04 μm wide (5). The carbon dioxide laser is by far the most common gas laser in the infrared region, emitting on any of several dozen lines in the 10.72 to 10.17 μm and 9.66 to 9.21 μm ranges. The two most common CO₂ lasers are the longitudinally excited glow-discharge laser and the transversely excited atmospheric (TEA) laser. The former is an efficient continuous-output laser, producing about 80 W of gain per meter of tube length; the latter produces short (typically 100 ns) pulses at instantaneous powers up to the megawatt level. The CO₂ laser is representative of a group of molecular gas lasers. The carbon monoxide laser produces about one-fifth the power of the CO₂ laser and functions on many lines between 4.9 and 8.3 μm. The nitrous oxide[10024-97-2], N₂O, laser has one-seventh the power of CO₂ and lases on about 100 lines between 10.3 and 11.1 μm. There are also many far-infrared molecular gas lasers that operate on rotational transitions between 40 and 2000 μm at powers up to 1 W. Most are optically pumped using CO₂ or N₂O lasers, and different wavelengths are produced by changing the gas used (6).

Infrared Materials and Optics. There are many metal halides and chalcogenides that are transparent over large portions of the infrared spectrum. Glass (qv), however, is limited to wavelengths shorter than 2.7 μm because of a strong absorption at 2.8 μm. The transmission of other oxides ends partway through the mid-infrared region, for example, at 3.5 μm for infrared-grade fused silica and at 5 μm for sapphire. The alkali halides are the most common mid-infrared window materials. A reststrahlen band governs the long-wavelength transmission limit, which shifts to longer wavelengths as the masses of the ions in the salt increase, ranging from 6 μm for lithium fluoride [7789-24-4] to 45 μm for cesium iodide [7789-17-5] The most popular for mid-infrared use are sodium chloride [7647-14-5] potassium chloride [7447-40-7] and potassium bromide [7758-02-3], which have cutoffs at 16 μm, 20 μm, and 25 μm, respectively. The alkali halides have excellent optical properties and are inexpensive, but are easily scratched, easily cleaved, and moisture sensitive. Other metal halides used as infrared windows are generally stronger and less moisture sensitive, but more expensive. Nonoxide chalcogenides and semiconductors are hard and some are quite inert. These generally have high indexes of refraction, ranging among commonly used materials from 2.25 for zinc sulfide [1314-98-3] to 4.02 for germanium. This produces large reflectivities unless antireflection coatings are applied. These materials are used when reflection is desired, as in laser output couplers. Zinc selenide [1315-09-9] is popular for attenuated total reflection because of its combination of chemical resistance, low absorptivity out to 18 μm, and high (2.43) refractive index.

At far infrared wavelengths polyethylene and poly(ethylene terephthalate) (PET or Mylar) are transparent and can be used for windows. Lenses are less commonly used in the ir region than in the visible and uv ranges, because ir lens materials tend to be fragile, expensive, or reflective. Also important is the greater amount of chromatic aberration encountered in the infrared region because of the wider wavelength range. The visible spectrum consists of a single octave of frequencies. In contrast, the near infrared spectrum encompasses roughly two octaves, and the mid-infrared and near infrared each encompass three. Lenses are used where the more compact, straight-line designs possible with lenses are required, such as in thermal imagers (7).

The reflectance of metals is generally higher in the infrared range than in the visible. Gold is the most widely used reflector when high reflectance is important. Its reflectance is above 98% throughout the infrared spectrum and above 99% for wavelengths longer than 1.5 μm. Its reflectance is low for wavelengths smaller than 0.6 μm, so gold is also desirable when visible light rejection is needed. Gold is soft but does not tarnish. It may be used bare if it is handled and cleaned carefully. The reflectance of aluminum dips at 0.82 μm, which is above 97% for wavelengths longer than 1.5 μm, making aluminum an economical choice. Aluminum oxidizes, so it is often given a protective coating. Magnesium fluoride [7783-40-6], MgF₂, and SiO are common protective films. Silver has a reflectance higher than gold in the near infrared region, but it is so vulnerable to oxidation that it is principally used for second surface mirrors. Metal-substrate mirrors are used if heat dissipation is important. Copper is popular both as a substrate for gold-coated mirrors and as a metal mirror itself. A fresh copper surface has a reflectance close to that of gold, but it is susceptible to tarnish. Dielectric coatings can be put on any of these metals to produce the highest possible reflectance over a limited range of wavelengths.

Colored absorption filters, which are commonly used in the visible spectrum, are used as long-wave pass filters in the near infrared region, where their component materials can transmit. Many semiconductors have a sharp transmission onset and can be used as long-wave pass filters in the near and mid-infrared regions. The transmission onset corresponds to the energy bandgap of the semiconductor, which can be adjusted by doping, as is done for extrinsic infrared detectors. The most commonly used infrared filter is the multilayer interference filter, which consists of a series of thin layers built up on a transparent substrate. The layers in the series alternate between high and low refractive index, so the infrared beam undergoes repeated reflection off the many parallel interfaces. The thicknesses and indexes of refraction of the layers are such that for the desired wavelengths backward reflections destructively interfere and cancel out, whereas forward reflections constructively interfere. A summary of commercially available short-wave pass, long-wave pass, and bandpass interference filters is available (7). Many bandpass interference filters also transmit at integer multiples of the nominal pass band. Unless these are deposited on substrates that are opaque at those longer wavelengths, additional long-wave attenuation may be needed. Interference filters are usually designed for use at normal beam incidence. The angle of incidence affects the pass band, because the effective thickness of the filter layers changes with angle. The angle of incidence has little effect on the bandwidth of a bandpass filter, but it shifts the center wavelength to shorter wavelengths. The center wavelength can be tuned over a range from 100 to ~93% of its normal incidence value by tilting it. At non-normal incidence an interference filter becomes polarizing. The bandpass is also temperature sensitive, shifting linearly to longer wavelengths with increasing temperature. A typical wavelength shift is 0.007% °C. Bandpass filters in which the layer thicknesses change smoothly with position are called variable bandpass filters. The layer thicknesses, and therefore the center wavelength of the pass band, are linear functions of angular position on a circular filter or of linear position on a rectangular filter. Interference films are also used on windows and lenses as antireflection coatings and on mirrors to maximize reflectivity.

Dichroic polarizers, the most common type in the visible spectrum, contain linear dye molecules oriented in one direction so that they preferentially absorb light of one polarization. Dichroic filters can produce a beam of good polarization purity, and are inexpensive, but they can only be used for low power applications. Dichroic filters are available out to roughly 2 μm, which is the absorption limit for suitable dyes. A birefringent prism splits an incident beam into two orthogonally polarized beams, which travel through the crystal at different angles. Various prism shapes (eg, Glan-Thompson, Wollaston) have been designed which differ principally in the way they separate the two beams. Birefringent prisms can handle higher powers than dichroic polarizers, but birefringent prisms are very orientation sensitive and expensive. Birefringent calcite prisms are available out to 2.3 μm. The simplest type of infrared polarizer, often referred to as a stack-of-plates polarizer, consists of a series of transparent, high refractive index plates positioned in the beam at Brewster's angle. At Brewster's angle, which is tan⁻¹ (n₂ / n₁) for plates of index n₂ in a medium of index n₁, only one polarization is reflected. The polarization purity of the remaining beam depends on the number of plates used. Stack-of-plates polarizers can handle high powers, but are sensitive to orientation. Interference filters are essentially stacks of plates, so they can be used as polarizers. The wire grid polarizer is unique to the infrared region and longer wavelengths. If parallel wires are positioned normal to an infrared beam and the spacing between wires is small compared to the wavelength, the wires strongly reflect the polarization having its K vector parallel to the wires but do not affect the other polarization. Wire grid polarizers, available for wavelengths longer than 1 μm, are compact, insensitive to angle, and expensive.

Optical Fibers. Transmission within an optical fiber relies on the total internal reflection of the electromagnetic wave as it travels in the fiber core (see FIBER OPTICS). The simplest type, a stepped-index fiber, consists of a uniform cylindrical core having an index of refraction n₁, covered by a cladding material having a lower index of refraction, n₂. A wave striking the boundary between the core and cladding is totally reflected if the angle between the boundary and the wave does not exceed the critical angle, π_c, defined as cosφ_c = n₂ / n₁. This limits the numerical aperture (NA) of the fiber:

$$NA = \sin\theta = n_1 \sin\theta_c$$

(1)

where θ is the maximum allowable angle between the axis of the fiber and the wave incident on the fiber end. The number of geometrical modes (paths) by which waves can be guided down a fiber is proportional to both the radius of the core and NA^2 . A greater number of modes means a greater amount of light may be launched into the fiber, but the higher order modes, which undergo more reflections, are prone to greater losses. Graded-index fibers, in which the refractive index increases smoothly with distance from the center of the fiber, more efficiently transmit fewer modes than stepped-index fibers of the same size. In cases where the synchrony of all waves entering a fiber must be maintained, a narrow fiber capable of sustaining only one path (a single-mode fiber) is required (8). Fiber attenuations are usually given in dB km or dB m. The attenuation value in dB per length is 10 times the absorbance value for the same length; thus an extinction coefficient of $\alpha \text{ cm}^{-1}$ equals an attenuation of $1000 \alpha \text{ dB m}$.

Silica-based glass fiber, used in the telecommunications industry, is the most mature of the fiber technologies. Silica fiber has achieved an attenuation of 0.2 dB km at $1.55 \text{ }\mu\text{m}$, its theoretical limit. The fiber is strong and inexpensive. Silica fibers can transmit well from 0.25 to $2 \text{ }\mu\text{m}$, but different formulations are used for the uv and ir ends of that range. For somewhat longer wavelengths, fluoride glasses are the preferred materials, and the fluorozirconates are the best developed of these (see FLUORINE COMPOUNDS, INORGANIC). Most fluoride glasses are a mixture of several fluorides, so they have become known by acronyms based on the first letters of the chemical symbols for the constituent metal elements. For example, ZBLAN fiber contains on a mol % basis 55.8° ZrF_4 , 14.4° BaF_2 , 5.8° LaF_3 , 3.8° AlF_3 , and 20.2° NaF . Fiber composition governs the transmission range. ZBLAN, for example, transmits down to $3.8 \text{ }\mu\text{m}$, but other fluorozirconates can reach $5.5 \text{ }\mu\text{m}$, and some $\text{BaF}_2\text{--ThF}_4$ glasses reach $7 \text{ }\mu\text{m}$ (9). The fluoride glasses are not as strong as the silica glasses and are moisture sensitive. State-of-the-art fluoride fibers have minimum losses below 1 dB km , but commercial fibers have losses of a few tens of dB km. Chalcogenide fibers are required for even longer wavelengths. Most chalcogenide fibers are nonstoichiometric mixtures of sulfur, selenium, or tellurium with one Group 15 (V) element or one Group 14 (IV) element, or both. Chalcogenide fibers are not as strong as fluoride fibers, but are not moisture sensitive. The transmission ranges of some chalcogenides extend to about $16 \text{ }\mu\text{m}$. State-of-the-art fibers have minimum losses of a few tens of dB km, but commercial ones have minimum losses of about 1 dB m . The principal source of attenuation in chalcogenide fibers is absorption by impurities, particularly oxygen, hydrogen, and water. As a result, attenuation in a chalcogenide fiber is highly wavelength dependent, showing absorption peaks from the various impurities (8).

Detectors. A great variety of devices has been used as infrared detectors, ranging from d-c thermocouples, used in the earliest commercial spectrometers, to the Evaporograph, in which an infrared image was focused on a thin film of oil. Evaporation of the oil produced colored interference fringes, which were then photographed; thus this was the earliest color thermal imaging system (see THERMOGRAPHY). Modern infrared detectors fall into two classes, thermal and quantum (or photon). A thermal detector has a blackened surface that absorbs incident radiation at all wavelengths and degrades it to heat. The resulting temperature change in the sensor element produces the detector signal. In quantum detectors the infrared absorption excites electrons, altering an electrical property of the detector, which is measured (see PHOTODETECTORS). Thermal detectors have sensitivities that are independent of wavelength, but are slow because a temperature change must occur. Quantum detectors are generally faster and more sensitive, but have a sensitivity that rises smoothly with increasing wavelength up to a long-wavelength limit beyond which it drops rapidly.

The responsivity, the noise equivalent power, and the specific detectivity are quantities often used to judge detector performance. The responsivity is defined as the root-mean-square (rms) signal voltage divided by the rms incident power producing that signal. The noise level is usually given as the noise equivalent power (NEP), the incident infrared power that would produce a signal equal to the rms noise signal of the detector:

$$\text{NEP} = \frac{V_n}{R_v} \tag{2}$$

where R_v is the responsivity in V W^{-1} and V_n is the rms signal produced by the noise, in units of $\text{V Hz}^{1/2}$. The root frequency unit arises from the fact that the voltage depends on the square root of the bandwidth of frequencies over which the measurement is made, so the measured voltage is divided by the square root of the bandwidth. The most commonly quoted performance specification of detectors is D^* , the specific detectivity. For a detector of area A ,

$$D^* = \frac{A}{\text{NEP}} \tag{3}$$

Because NEP is roughly proportional to $A^{1/2}$, D^* is more useful for comparing detectors of differing sizes. D^* depends on the wavelength distribution striking the detector (if it is quantum) and the frequency at which the radiation is modulated, so these measurement parameters need to be included for a D^* value to have meaning. Often detectivity is written as $D^*(T,f,b)$, where T is the temperature of the blackbody source of radiation or the wavelength of the monochromatic radiation, f is the modulation frequency, and b is the bandwidth of the measurement (10,11).

Only a few types of thermal detectors are in common use. Thermocouples and thermopiles (groups of thermocouples connected in series) use the thermal emf of a junction to measure radiation. Thermocouples are slow and not very sensitive, but robust and inexpensive. Their principal use is in energy measurement, such as laser power meters. The pyroelectric detector is by far the most commonly used type of thermal detector. Its thermal sensor is a spontaneously polarized ferroelectric material, deuterated triglycine sulfate (DTGS) and lithium tantalate [12031-66-2] being the most popular (see FERROELECTRICS). A thin sheet of the ferroelectric is sandwiched between two electrodes, forming a capacitor. A temperature change alters the polarization, which induces an external current flow until the charge distribution is equilibrated. Pyroelectric detectors are therefore inherently a-c devices. Although faster than most other thermal detectors, the responsivity of pyroelectric detectors is roughly proportional to the inverse of the modulation frequency, down to around 1 Hz, below which the responsivity drops. Typical maximum D^* values for high quality pyroelectric detectors are around $10^8 \text{ cm}^2 \cdot \text{Hz}^{1/2} \text{ W}^{-1}$. Bolometers measure the change in conductance of a temperature sensitive material. Room temperature bolometers, based on metal or semiconductor (thermistor) sensors, are slow and have modest sensitivity. By contrast, liquid-helium cooled germanium and hot-electron bolometers are both faster and more sensitive. The germanium bolometer operates much like room temperature bolometers, but the active cooling ensures fast thermal response and greatly reduces noise, thereby raising detectivity. It is generally used in the 400 to 5 cm^{-1} region. The hot-electron bolometer is based on indium antimonide, InSb, that has been doped so that it contains free carrier electrons, which can directly absorb the incident radiation. This leads to a fast, sensitive thermal detector, but it is limited to energies below about 50 cm^{-1} because of the absorption spectrum of the electrons (12,13).

In quantum detectors, electrons are directly excited to higher energy levels by the absorption of photons. All quantum detectors have a maximum wavelength (minimum wavenumber) beyond which they do not function, which is defined by the size of the energy gap the electrons must jump. In photoemissive detectors, such as phototubes and photomultipliers, the electrons acquire sufficient energy to escape the surface of the material. Photoemission requires an electron energy of more than 1 eV; thus such detectors are limited to wavelengths of roughly $1 \text{ }\mu\text{m}$ or shorter. Most other infrared quantum detectors use photosensitive semiconductors as sensors (qv). If the electronic excitation involves only energy levels of the pure semiconductor material, then the device is an intrinsic detector, but the available energy levels can be tailored by doping in an impurity element, producing an extrinsic detector. In photoconductive detectors the electronic excitation produces an increase in the conductivity of the semiconductor, which is measured via an external bias (see SEMICONDUCTORS). Both extrinsic and intrinsic photoconductive detectors are common. Photovoltaic detectors, commonly called photodiodes, produce a voltage pulse whenever photons are absorbed, so they do not require an external bias. Photodiodes, which are usually intrinsic, consist of a junction between n-type and p-type semiconductors, as in ordinary diodes. The electronic excitation produces an electron-hole pair, which is pulled apart by the intrinsic potential of the junction, producing the external voltage. Avalanche photodiodes have a large reverse bias applied to them, which results in signal amplification and an improvement in the signal-to-noise ratio. Table 1 lists the more common semiconductors used in detectors. As the table shows, D^* virtually always increases as the operating temperature decreases, but the cutoff wavelength can shift up or down. $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ is an example of a ternary detector, in which the value of x controls the cutoff wavelength. Photoconductive detectors are generally simpler to couple to low noise amplifiers; photodiodes generally have lower power consumption because these have no external bias, and better high frequency performance (15,16).

Table 1. Characteristics of Selected Quantum Detectors^a

Material	Temperature, °C	Cutoff wavelength, μm	cm·Hz W ^b	Frequency for D*, Hz
<i>Photovoltaic^c</i>				
InGaAs	25°C	1.65	1 × 10 ¹²	300
Ge	196°C	1.6	6 × 10 ¹³	300
	25°C	1.8	3 × 10 ¹¹	300
InAs	196°C	3.1	6 × 10 ¹¹	1,200
InSb	196°C	5.5	1 × 10 ¹¹	1,200
<i>Photoconductive</i>				
PbS	25°C	2.9	1 × 10 ¹¹	600
	77°C	3.6	1 × 10 ¹²	600
PbSe	25°C	4.8	1 × 10 ⁹	600
	77°C	5.8	6 × 10 ⁹	600
Hg _x Cd _{1-x} Te ^d	196°C	13.5	5 × 10 ¹⁰	10,000
	196°C	16.6	4 × 10 ¹⁰	10,000
	196°C	22	1 × 10 ¹⁰	10,000
Ge:Cu doped	269°C	32	5 × 10 ⁹	1,000
Ge:Zn doped	269°C	42	1 × 10 ¹⁰	1,000

^a Ref. 14.

^b Measured at the wavelength of maximum detectivity and the frequency given with a bandwidth of 1 Hz.

^c Commonly called photodiodes.

^d For different values of x.

Infrared photographic film can be used for imaging out to 1.2 μm, the wavelength limit for available emulsions. The pyroelectric vidicon (an infrared-sensitive television tube) is still used, but most infrared imaging is now done with solid-state detector arrays, although many systems combine arrays with image scanning. Pyroelectric arrays are the principal thermal type produced, but most are quantum. A detector array consists of two components, the actual array of detector elements and an array of signal processing units. Commonly, the processing array is a charge-coupled device (CCD), which receives the photon-induced charges from the detector elements and transfers them to the measuring electronics. The CCD may keep the charge from an individual detector element separate from all others or it may combine the charges from elements in such a manner that the image integrity is maintained. A detector array can be either monolithic, in which the detector and signal processor elements are all made on the same semiconductor crystal, or hybrid, in which the detectors and signal processors are made on different materials and then bonded together (17). Monolithic silicon arrays, used in the visible range, can extend out to 1 μm. Hybrid InSb, hybrid lead selenide [12069-00-0], PbSe, and monolithic platinum silicide [12137-83-6], PtSi, dominate the 1 to 5 μm range. At longer wavelengths arrays are rarely used, but mercury cadmium telluride [29870-72-2], HgCdTe, is available. Array development is a particularly active field and new materials are reaching the market.

Radiometers and Thermal Imagers. A radiometer is essentially a calibrated detector and its supporting electronics used to measure the radiance of a source. Numerous accessories can be added to this basic description. Radiometers can include a lens or mirror system to define the area monitored. Inclusion of a beamsplitter and eyepiece allows the operator to view the monitored area and aim the instrument. Integral reference sources are used for calibration. An infrared thermometer is a radiometer that incorporates a filter to limit the measurement to a narrow wavelength range. The radiance measurement is then used to calculate the surface temperature of the radiant object. Which wavelength band is best depends on the temperature range of the radiant object and its composition. A two-color infrared thermometer makes measurements through two filters at different wavelengths then derives the temperature from their ratio. This is inherently more reliable than the absolute measurement at a single wavelength because the derived temperature is independent of both the emissivity of the object and the presence of dust or other scattering contamination in the infrared path.

Thermal imagers are used for radiometry, spectroradiometry, and thermal imaging. Several parameters are used to define the performance of thermal imagers. The noise equivalent temperature difference (NETD or NEΔT) is analogous to the NEP for detectors. It is the temperature difference that produces a signal-to-noise ratio of 1 and is therefore just detectable. The NETD is proportional to the NEP of the detector in the imager. NETDs of 0.1°C or less are commonly available. The instantaneous field of view (IFOV) is a limit on imager resolution:

$$\text{IFOV} = A (F_N D)$$

(4)

where A is the linear dimension of the detector, F_N is the f -number of the detector optics, and D is the diameter of the entrance pupil. The modulation transfer function (MTF) is an empirical measure of how well an image is reproduced. If a sinusoidal intensity pattern incident on the imager has a contrast of x and the imager reproduces the pattern with a contrast of y , then $\text{MTF} = y/x$ at that spatial frequency. The minimum resolvable temperature difference (MRTD) is a parameter that combines the NETD and the MTF. It is the minimum temperature difference between a standard four-bar pattern and its background that allows the bars to be perceived (15).

The conceptually simplest imager uses a staring array and is an analogue of the photographic camera. The complete image is focused onto a detector array and each pixel in the picture corresponds to the signal from a specific detector in the array. A staring-array arrangement gets the best thermal sensitivity but the worst spatial resolution from a particular detector array. When a staring pyroelectric array (or vidicon) is used, a chopper is usually built in because a pyroelectric detector is an inherently a-c device. Detector technology often cannot provide adequate resolution by this method and some form of image scanning must be used. The scanning method may be serial, parallel, or some variation or combination of these. In their basic forms, both use a linear detector array. In parallel scanning, the linear array is lined up along one dimension of the image and the image is scanned over the array in the other dimension. Each detector element is responsible for one stripe within the final picture. This method tends to produce a liney image, but the scanning is mechanically simple and can be performed at a low rate, so the electronic bandwidth need not be large. In serial scanning, the image is raster scanned over the detector with the array aligned with the raster lines. Each detector element in the array views the full image, but at a slightly different point in time from its neighbors. The image is constructed by adding together the signals from the individual elements after the appropriate time delays have been added, called time delay and integration (TDI). Serial scanning gives a uniform image, but requires a higher electronic bandwidth and more complicated scanning optics. Serial-parallel scanning combines the two modes with a rectangular detector array. The image is raster scanned over the array, but the separation between lines in the raster pattern is equal to the dimension of the array, so each stripe in the final picture is observed by a specific row of elements in the array. A signal processing in the element (SPRITE) detector has TDI built in. A bias applied to the array causes the photon-generated charges to drift through the array at the same speed as the image scanning, so the signal generated by each element is synchronized with the passing stream of charges (15). All of the technology developed for the electronic storage, transmission, and analysis of images can be used with thermal imagers. The data can be presented as black-and-white or false-color images using user-selected radiometric or temperature scales. Precise measurements of the temperature can be derived for points or lines within the image, or the average temperature within a section of the image. Some include internal reference sources for temperature calibration. Various models provide outputs for computers and videocassette recorders.

Grating Spectrometers. Spectrometers can be classified as spectrophotometers and spectroradiometers. Spectrophotometers are intended for testing a sample by measuring its interaction with radiation, and have a built-in source. Spectroradiometers analyze the emission from a test object; thus the sample is the source. Infrared grating spectrometers are similar to their uv visible counterparts (see SPECTROSCOPY). The infrared radiation enters through a slit, its wavelengths are dispersed by a grating, then the dispersed wavelengths either strike a detector array or pass one at a time through an exit slit as the grating rotates. The law for the grating dispersion is

$$d(\sin\theta_i \pm \sin\theta_d) = m\lambda \tag{5}$$

where d is the spacing between grating lines, θ_i and θ_d are the angles of the incident and diffracted beams relative to the grating normal, λ is the wavelength, and m is the grating order (an integer). The infrared spectrum spans a much wider range of wavelengths than the visible and uv, which complicates spectrometer design, but ir gratings require fewer lines per millimeter to achieve the same wavelength resolution. Grating efficiency, quantum detector sensitivity, and source output all vary appreciably with wavelength; thus grating spectrometers must compensate for these changes as they scan. In addition, grating orders overlap if the radiation source emits over more than a small fraction of the infrared range. Infrared spectrometers often include various blocking (bandpass) filters to prevent order overlap, multiple interchangeable slits to compensate for source output changes, more than one grating, and more than one detector. The instrument automatically switches among the various slits, filters, gratings, and detectors as it scans so as to provide the optimum combination in each wavelength range. Many infrared spectrometer systems also include a beam chopper and a phase-sensitive detector. In double-beam spectrophotometers, the chopper is reflective and arranged so that the radiation is alternately sent along the sample and reference beam paths; then the two beams are recombined at the detector. The a-c component of the detector signal is null unless some excess loss, presumably absorption, has occurred in the sample path.

Fourier-Transform Infrared Spectrometers. Most Fourier-transform infrared spectrometers are based on an interferometer designed by Michelson in 1891. Figure 1 shows a basic structure. A collimated beam of infrared radiation is directed into an interferometer, consisting of a beamsplitter, a fixed mirror, and a moving mirror. In most spectrometers the moving mirror glides at constant velocity toward and away from the beamsplitter. The beamsplitter reflects part of the incoming beam to one mirror and passes the rest on to the other. The mirrors reflect the beams back on themselves. At the beamsplitter the returning beams again partly reflect and partly transmit, so the beams recombine and mutually interfere. The recombined beam is focused into the sample compartment. The difference in path lengths followed by the beams reflected off the fixed and moving mirrors, called the retardation, δ , changes as the moving mirror glides back and forth, creating a temporal interference pattern. To understand the operation of the interferometer, consider radiation at a single wavelength, λ . When the moving and fixed mirrors are equidistant from the beamsplitter, ie, at $\delta = 0$, the split infrared beams follow equal paths and are exactly in phase when they recombine at the beamsplitter, so they constructively interfere and the illumination of the detector is at a maximum. As the retardation increases, the split beams recombine increasingly out of phase. When the moving mirror has shifted by $j\lambda$, the retardation is $j\lambda$, the recombining beams are 180° out of phase, they destructively interfere, and the detector illumination is at a minimum. When the retardation reaches λ , the beams are back in step and constructively interfere. As the moving mirror travels at a velocity v , the retardation changes at a rate $2v$ and interference produces an a-c detector signal consisting of a sine wave of frequency $2v/\lambda$. With a wideband source, each wavelength produces its own sinusoidal intensity variation, and the detector signal, called the interferogram, is the sum of all of them. Ftir spectrometers contain a low power laser and a laser detector. The laser beam passes through the interferometer, so the beam intensity has a sinusoidal variation at the laser detector. The spectrometer monitors this variation to determine the position of the moving mirror and to trigger digitization of the infrared-detector signal. Most spectrometers sample the signal every time the retardation changes by one or one-half laser wavelength (18,19).

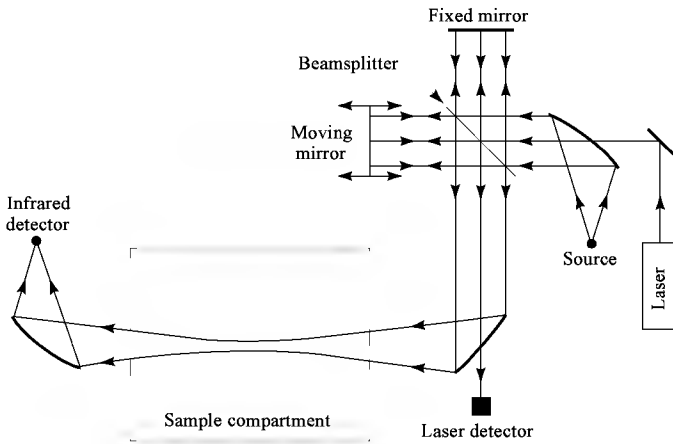


Fig. 1. Schematic of the optical layout of a Fourier-transform spectrometer.

The interferogram and the spectrum are related by the Fourier-transform pair:

$$I(\delta) = \int_{-\infty}^{+\infty} B(\tilde{\nu})e^{i2\pi\tilde{\nu}\delta} d\tilde{\nu} \qquad B(\tilde{\nu}) = \int_{-\infty}^{+\infty} I(\delta)e^{-i2\pi\tilde{\nu}\delta} d\delta \tag{6}$$

where $I(\delta)$ is the interferogram and $B(\tilde{\nu})$ is the spectrum, which is a function of the wavenumber $\tilde{\nu}$ ($\tilde{\nu} = 1/\lambda$). The fast Fourier-transform algorithm calculates the spectrum from the digitized interferogram in real time. The fast Fourier transform acts on an interferogram that is discrete and finite in length, unlike the continuous and infinite integrations shown above. The spacing of data points in the interferogram limits the valid wavenumber range of the resulting spectrum. The Nyquist criterion on sampling waveforms states that any sum of sinusoids can be accurately represented by sampling at a frequency that is at least twice the highest frequency present in the sum. For a Fourier-transform spectrometer, this means the interferogram must be sampled at a rate of $2\tilde{\nu}_{\text{max}}$ or greater if $\tilde{\nu}_{\text{max}}$ is the highest wavenumber observed by the detector. If higher frequencies are present they undergo aliasing or folding and appear at false positions in the spectrum. The limited extent of the interferogram limits the resolution of the resulting spectrum. A resolution of $\Delta\tilde{\nu}$ wavenumbers requires an interferogram that extends out to about a retardation of $(\Delta\tilde{\nu})^{-1}$. The connection between resolution and interferogram length can be understood by considering how to resolve two wavenumbers, $\tilde{\nu}_1$ and $\tilde{\nu}_2$, which differ from one another by a small value, $\Delta\tilde{\nu} = \tilde{\nu}_1 - \tilde{\nu}_2$. As the retardation goes from zero to $(\Delta\tilde{\nu})^{-1}$, the $\tilde{\nu}_2$ radiation undergoes exactly one less intensity modulation than the $\tilde{\nu}_1$ radiation, so the two can be differentiated. In practice, the resolution also depends on apodization. Apodization is the multiplying of the interferogram, prior to transformation, by a function that makes the interferogram taper smoothly down to zero intensity at its ends. Abruptly truncating the infinite interferogram at some maximum retardation produces an artifactual series of small peaks surrounding each real peak. Apodization reduces or removes the side peaks but increases the width of the real peak (ie, decreases resolution) (18,20).

Fourier-transform spectrometers have two advantages relative to monochromators. These are the multiplex or Fellgett's advantage and the throughput or Jacquinot's advantage. The former is the improvement in signal-to-noise ratio that results because an interferometer is measuring all parts of a spectrum at the same time; the latter is the much larger amount of radiation admitted through the limiting (typically circular) aperture of an interferometer

than through the narrow slit of a monochromator. There are, however, certain factors that can defeat these advantages. Fluctuations in source intensity, detector sensitivity, or throughput diminish the multiplex advantage. Because of multiplexing, fluctuation noise in any part of the spectrum affects the measurement of all parts of the spectrum. The throughput and multiplex advantages are important only if the noise level is independent of the signal level. This is true of most far and mid-infrared detectors, so the Fourier-transform approach now dominates in these regions. When the noise level is proportional to the square root of the signal level, as in the shot noise of a photomultiplier tube, there is no multiplex advantage. If noise is even more sensitive to signal level, there is actually a multiplex disadvantage (18,21).

Interference Filter Spectrometers. Spectrometers based on bandpass interference filters can be used wherever the higher resolution afforded by grating and interferometer-based instruments is not needed, so these are particularly popular for near infrared applications. Filter spectrometers are generally smaller and more rugged than grating and interferometer instruments, and are particularly useful as industrial and portable instruments. The simplest filter instrument uses a series of bandpass filters mounted on a wheel in front of a detector. The filters are sequentially positioned in front of the detector and a measurement is made through each. The advantages of this design are its wavelength reproducibility and the ease with which calibrations can be transferred among instruments. The disadvantage is its lack of flexibility. Any possible source of interference must be identified and compensation made. Filter wheel instruments are best used for monitoring tasks in well-defined situations, where reliability is a strong advantage and lack of flexibility is not important. A variable bandpass filter spectrometer combines a circular variable filter with a slit input aperture. The one disadvantage of such an instrument is its low resolution. The effective bandpass of the instrument is greater than the intrinsic bandpass of the filter, because the center wavelength of the filter bandpass varies across the width of the slit. The tilting filter spectrometer has a bandpass equal to that of its filters, but it is not as compact as the variable filter instrument. The center wavelength of a bandpass filter can be tuned over roughly a 7% range by tilting the filter in the infrared beam. In a tilting filter spectrometer a series of bandpass filters are edge-mounted on the face of a rotating disk so that as the disk rotates each filter is successively brought into the infrared beam and tilted. The full spectrum is built up piecewise from the separate filters.

Acousto-optic Filters. The newest type of spectrometer to become commercially available is the acousto-optic tunable filter (AOTF). An AOTF is a solid-state, electronically tunable bandpass filter based on the diffraction of optical waves by acoustic waves in an optically anisotropic crystal. An AOTF device consists of a birefringent crystal to which a piezoelectric transducer has been bonded (see **PIEZOELECTRICS**). When the transducer is driven by a radio frequency signal, it launches acoustic waves into the crystal that produce a periodic modulation of the index of refraction, and thereby constitute a moving grating which diffracts only those optical frequencies having the proper phase relationship. An AOTF has a number of both novel and advantageous properties. It can scan very rapidly, requiring only tens of microseconds to switch between wavelengths. Driving it using multiple radio frequency signals tunes it to more than one wavelength simultaneously. It has a large (up to 10 mm × 10 mm in commercial devices) aperture. Diffraction efficiencies as high as 90% have been achieved. The phase relationship required for diffraction can be achieved in collinear and noncollinear configurations, both of which are illustrated in Figure 2. In the collinear form, the longitudinal acoustic wave produced by the transducer is reflected off the end of the crystal to produce a shear wave, which propagates collinearly with the polarized optical beam. The optical wave in phase with the acoustic wave has its polarization axis rotated by 90°. Symmetry prevents construction of collinear devices from certain crystals, so the noncollinear configuration is more common. In the noncollinear AOTF, the incident light is diffracted into two beams of mutually orthogonal polarization that exit the crystal at different angles. One of the diffracted beams can be selected for use with an aperture or a polarizer. The collinear configuration has the longer interaction length between the optical and acoustic waves, and thus it has the better spectral resolution, but the noncollinear configuration has the larger aperture. AOTFs have modest resolution. Peaks in an AOTF-produced spectrum have small side peaks, as in an unapodized Fourier spectrum. At present commercially available infrared AOTFs are all of the noncollinear form. *para*-Tellurite, TeO₂, is perhaps the best material for the visible and near infrared regions. It has a high acoustic figure of merit and can be used from 0.34 to 4.5 μm. Thallium arsenic selenide has been commercially used for devices between 2.5 and 17 μm, but there are other crystals available for the mid-infrared region (22).

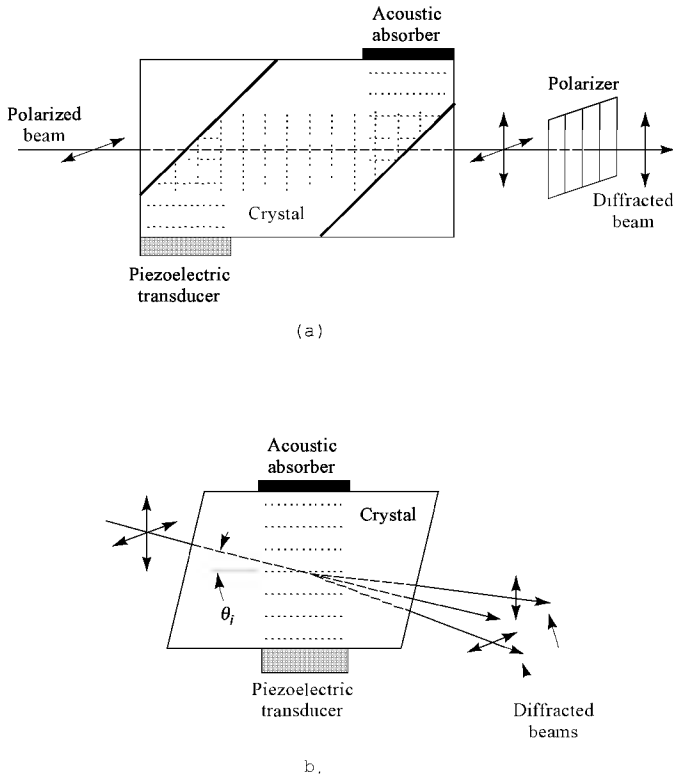


Fig. 2. Structures of acousto-optic tunable filters: (a) collinear and (b) noncollinear.

Spectrometry

The range of photon energies (160 to 0.12 kJ mol (38–0.03 kcal mol)) within the infrared region corresponds to the energies of vibrational and rotational transitions of individual molecules, of electronic transitions in many semiconductors, and of vibrational transitions in crystalline lattices. Semiconductor electronics and crystal lattice transitions are beyond the scope of this article.

At low energies, the rotational and vibrational motions of molecules can be considered separately. The simplest model for rotational energy levels is the rigid dumbbell with quantized angular momentum. It has a series of rotational levels having energy

$$E_{\text{rot}} = BJ(J + 1)$$

(7)

where *B* is the rotational constant and *J* is the rotational quantum number, which must be an integer. Only transitions in which *J* changes by 1, ie,

$\Delta J = \pm 1$, can be induced by absorbing or emitting a photon, so the rotational spectrum consists of a series of equally spaced lines at photon energies of $2B\bar{J}'$, where $\bar{J}'$ refers to the upper level involved in the transition. Many diatomic molecules have rotational spectra that come quite close to this idealized spectrum at low rotational energies. Larger molecules can rotate around three mutually orthogonal axes, and expressions for the energy levels become more complex as the molecules become less symmetric. Nevertheless, many molecules have rotational spectra with roughly equal spacing between lines or groups of lines. Only transitions from populated levels, ie, levels up to roughly kT in energy, $\sim 200\text{ cm}^{-1}$ at 298 K, can be observed. Values for B range from 60 cm^{-1} down to less than 1 cm^{-1} , so rotational spectra occur in the far infrared and microwave regions (see also MICROWAVE TECHNOLOGY).

The simplest model for vibrational energy levels is the harmonic oscillator, which has allowed levels with energy

$$E_{\text{vib}} = v_0(v + 1/2)$$

(8)

where v_0 is the vibrational constant and v is the vibrational quantum number, which must be an integer. The selection rule for the harmonic oscillator is $\Delta v = \pm 1$, so all transitions occur at the same energy, v_0 . Many diatomic molecules are only slightly anharmonic. The anharmonicity weakens the selection rules, so absorptions with $\Delta v = \pm 2, \pm 3$ and so on, are observable but weak. Anharmonicity also changes the spacing between successive vibrational levels, so transitions with the same Δv no longer exactly overlap. In real molecules, vibrational and rotational transitions occur simultaneously. The selection rule $\Delta J = \pm 1$ is valid for vibrational transitions of diatomic molecules, and some diatomic molecules also have $\Delta J = 0$. The absorption lines owing to $\Delta J = -1, 0$ and $+1$ in a vibrational transition are called the P, Q, and R branches, respectively. The rotational constants of the upper and lower states are not equal, so the spacing grows with J in either the R or P branch and shrinks in the other branch. Almost all v_0 lie between 50 and 3500 cm^{-1} , so the dominant $\Delta v = \pm 1$ transitions fall in the far and mid-infrared ranges. The size of v_0 also means only one or a few vibrational levels are populated at ambient temperatures. The absorption or emission of an electromagnetic wave must change the electric dipole moment of a molecule, so only vibrational transitions that change the dipole moment are infrared active, ie, appear in infrared spectra. Similarly, only molecules with permanent electric dipoles can have far infrared spectra from purely rotational transitions. This means symmetric diatomic molecules, like H_2 and Cl_2 , have no infrared spectrum.

Different rules apply to Raman spectroscopy, so symmetric diatomic molecules do have Raman spectra (see INFRARED TECHNOLOGY AND RAMAN SPECTROSCOPY, RAMAN SPECTROSCOPY) (23,24).

Normal Modes, Group Frequencies, and Band Shapes. The number of allowed vibrational modes in a molecule is $3N - 6, 3N - 5$ for a planar molecule, where N is the number of atoms in the molecule. The allowed vibrational modes involve groups of bonds or all bonds in a molecule, and are called normal modes. The normal modes depend on the symmetry of the molecule and can be identified using the rules of group theory (23,25). Like the spectra of diatomic molecules, the vibrational spectra of large molecules are dominated by $\Delta v = \pm 1$ transitions, called fundamentals, but transitions having $\Delta v = \pm 2$ and larger, called overtones, and transitions in which the quantum numbers of two or more normal modes change simultaneously, called combination bands, may also occur. Whether these occur depends on their symmetry. For condensed-phase samples, overtone and combination bands are generally 10 to 1000 times weaker than fundamentals. Virtually all fundamentals fall in the far and mid-infrared regions, so the near infrared region encompasses only overtones and combination bands (and low lying electronic transitions).

Predicting the normal modes of most molecules is very difficult. Certain chemical bonds and functional groups usually produce an infrared peak near the same wavenumber location regardless of the molecule in which they occur. These characteristic band location ranges are called group frequencies, and are strong diagnostic tools for identifying an unknown sample. Extensive charts correlating bonds and functional groups with empirically determined group frequencies have been developed (24,26). Often some indication of the appearance of the spectrum band (eg, strong, moderate, broad) is also given. Group frequency data must be used carefully. Mechanical application of correlation charts usually does not give good results. Because of group frequency overlap, the absence of a band at the group frequency is stronger evidence than its presence. The region from 1400 to 600 cm^{-1} , called the fingerprint region, is usually very complicated in the spectra of organic molecules. Typically, the bands in this region are not readily assigned to a specific group, but the pattern of bands is characteristic of the specific molecule (24,27).

Nothing affects the appearance of the infrared spectrum of a substance more than its physical state. A gaseous molecule is free to rotate, so it has a purely rotational far infrared spectrum and mid-infrared vibrational transitions consisting of many sharp rotational lines. Although such densely structured spectra potentially provide the greatest amount of data, these are often more difficult to use analytically than condensed-phase spectra. Group frequency correlations are difficult for gas spectra, and gas-phase linewidths are dependent on the total pressure of the sample (not just the partial pressure of the analyte) via the phenomenon of pressure broadening. Increased pressure results in more frequent molecular collisions, which interferes with free rotation and widens absorption peaks. Pressure broadening makes absorption peaks appear to grow in size with total pressure unless the spectrum is recorded at high resolution. By contrast, the rotation of a molecule in the liquid state (neat or solution) is effectively suppressed. There is no purely rotational spectrum and the vibrational band shapes have a symmetric, statistical appearance that is largely independent of the molecule involved. To a good approximation, the band shape, A_{ν} , fits the Lorentz function:

$$A_{\nu} = A_{peak} \frac{\gamma^2}{\gamma^2 + (\bar{\nu} - \bar{\nu}_0)^2}$$

(9)

where A_{peak} and $\bar{\nu}_0$ are the absorbance and wavenumber, respectively, at the peak (center) of the band, $\bar{\nu}$ is the wavenumber, and γ is the half width of the band at half height. Liquid band positions are usually shifted slightly downward from vapor positions. Both band positions and widths of solute spectra are affected by solute–solvent interactions. Spectra of solid-phase samples are similar to those of liquids, but intermolecular interactions in solids can be nonisotropic. In spectra of crystalline samples, vibrational bands tend to be sharper and may split in two, and new bands may also appear. If polarized infrared radiation is used, both crystalline samples and stressed amorphous samples (such as a stretched polymer film) show directional effects (28,29).

Transmission, Absorption, and Beer's Law. The majority of infrared spectrometry is still done by the classic method of transmission spectrometry; the intensity of an infrared beam passing completely through a sample is measured. The standard description of how much radiation passes through the sample is that of Beer's law (or the Bouguer-Beer-Lambert law):

$$I = I_0 10^{-ac\ell}$$

(10)

where I_0 and I are the intensities of the infrared beam entering and exiting the sample, respectively, a and c are the absorption coefficient (or extinction coefficient or absorptivity) and the concentration of the absorbing species, and ℓ is the path length of the beam in the sample. The names for a are sometimes applied to the product ac as well. If more than one species absorbs at the wavelength of interest, then ac is the sum of contributions from the absorbers. Spectra are usually presented in terms of the absorbance, A , or transmittance, T , which are defined as follows:

$$T = \frac{I}{I_0} = 10^{-ac\ell} \quad A = -\log_{10} T = ac\ell$$

(11)

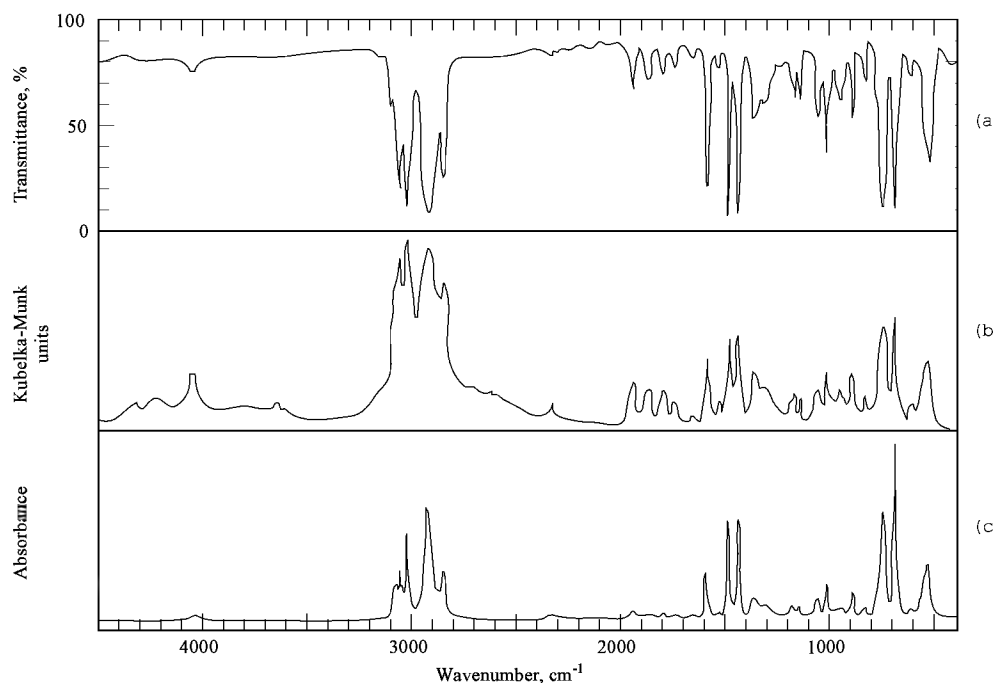
The linearity between A and c makes the concept of absorbance so useful that measurements made by sampling methods other than transmission are usually converted to a scale proportional to absorbance. The linearity between A and c is maintained only if the resolution of the spectrometer is adequate to eliminate contributions from wavelengths not absorbed by the species being measured. In addition, the apparent value of a is very dependent on resolution because a is a strong function of wavelength (30,31).

Near Infrared Spectrometry. A renaissance in near infrared spectrometry started in the late 1960s when the near infrared region began to be

used for the analysis of moisture, oil, and protein levels in grain and grain products (see *WHEAT AND OTHER CEREAL GRAINS*). The strength of near infrared analysis is its ability to handle condensed-phase samples with little or no sample preparation. The near infrared region is good both for samples that are hard to thin or dilute, like agricultural products, and for process-line monitoring, where analysis time must be kept to a minimum. This region contains only overtone and combination bands, which are generally broader and one to three orders of magnitude weaker than the fundamental bands. The weakness and overlapping of these broad bands were why the region was previously ignored, but the weakness is also the reason for its resurgence. The low absorption coefficients allow longer, more practical path lengths for transmission spectrometry and make reflection spectrometry of many undiluted, as-is samples practical. The ease in acquiring spectra in the near infrared region is balanced by the complexity of analysis. The weak, overlapping nature of the bands makes identification of a sample by visual inspection of its spectrum unlikely. Group frequencies have generally not been determined, with the exception of those for bonds to hydrogen, particularly C—H, N—H, and O—H (32). Overtone bands of these tend to be stronger and narrower than most near infrared features and often dominate a near infrared spectrum. The overlapping nature of the bands often prevents accurate quantitation based on peak height or area. Beer's law is valid, but it may be difficult to apply. Modern multivariate techniques are popular. Prior to these techniques, various ad hoc data treatments were developed that often had little theoretical basis, but were used because they worked. One data treatment that is very common in the near infrared region, but little used elsewhere, is the representation of the sample reflectance, R , in the form $\log_{10}(1/R)$. The practical reason for this is that $\log_{10}(1/R)$ is often an approximately linear function of concentration over ranges of interest. The more theoretically justifiable data treatment for reflectance, the Kubelka-Munk function, requires an accurate reflectance standard, which $\log_{10}(1/R)$ does not need. Both grating and *ftir* spectrometers function well in the near infrared range, but interference-filter spectrometers are especially popular. Their high throughput and low resolution complements the broad, weak nature of near infrared structure. Instruments with up to 20 fixed filters or seven tilting filters are available (33,34).

Sampling Methods. A wide variety of methods exists for the investigation of samples by infrared spectrometry. The choice depends on the nature of the sample, the kind of information desired, and the time available. Transmission, emission, and photoacoustic absorption are applicable to gases and vapors, but the reflection methods require a sample surface. When applied to condensed-phase samples, the methods have different levels of surface sensitivity and selectivity. Transmission and emission have no surface selectivity; these produce spectra of the sample bulk. The reflection techniques and photoacoustic absorption are surface selective; these can be used with infrared-opaque samples. Attenuated total reflection has the greatest surface selectivity, being limited to a sampling depth of roughly one-tenth of a wavelength. Specular reflection, diffuse reflection, and photoacoustic absorption all have penetration depths of up to a few tens of micrometers.

Transmission is the classic sampling method. It is highly reproducible, so it is accurately quantitative. Figure 3a shows the transmission spectrum of polystyrene film. Transmission often requires laborious preparation of solid samples, but it is still the principal sampling method for gases and liquids. A cylinder a few centimeters long with an infrared window at each end is a good transmission cell for most gases at partial pressures above 1 kPa (0.01 atm). Trace-gas analysis requires a long-path cell, which consists of a gas chamber having internal mirrors at the ends that reflect and refocus the infrared beam so that it passes through the chamber repeatedly. Commercial cells have path lengths of up to a few hundred meters and provide limits of detection down to parts per billion in an atmospheric pressure sample. Some cells have adjustable mirrors for changing the path length by changing the number of beam reflections. Because liquids are a thousand times more dense than gases, the optimum path lengths for liquids are a thousand times thinner. Quantitative work requires a cell having a fixed path length in the range of 0.1 to 5 mm. For qualitative work, demountable cells, in which an annular spacer defines the path length between two windows, provide path lengths down to 0.01 mm. Alternatively, the liquid may be squeezed to a thin film between two windows, or if it is not volatile it may be smeared on a window or on commercially available thin plastic films or wire grids. Most solids require preparation before they can be analyzed by transmission. The two classic preparation methods, making mulls and pellets, both involve grinding the sample and dispersing it at a concentration of 0.1 to 1% in a transparent matrix. The sample must be ground to a particle size smaller than the infrared wavelength involved to minimize scattering. A mull is prepared by grinding the sample in mineral oil (commonly called Nujol) or halocarbon oil. The suspension is then treated as a liquid sample. A pellet is made by grinding the sample, mixing it with a transparent powder (KBr most commonly), then pressing it into a disk, which is positioned in the beam path. The oils all absorb somewhere in both the near and mid-infrared ranges. The pellet method is more quantitative, and KBr is transparent, but hygroscopic, so water absorptions are common. Thin solid samples, such as plastic films, can be placed directly in the infrared beam, but if the film surfaces are smooth and parallel, the resulting spectrum may show interference fringes caused by repeated reflection of the beam at the sample surfaces (35,36).



spectrometry). Diffuse reflection is most commonly used on powders and other highly light scattering samples. The incident infrared radiation undergoes numerous reflections within the sample before escaping from the surface. The band shapes and data treatments differ between diffuse and specular reflection, so a diffuse reflection sampler is usually designed to avoid collecting the specularly reflected radiation. Samplers usually collect the radiation from as large a solid angle as practical (excluding the specular angle) so as to average over the sample and to maximize signal. The most common form of data treatment is based on the Kubelka-Munk model which relates the reflectance and the absorption coefficient, α

$$\frac{2.303ac}{s} = \frac{(1 - R_\infty)^2}{2R_\infty}$$

(12)

where c is the concentration of the absorbing species, s is the scattering coefficient of the sample, and R_∞ is the reflectance of an infinitely thick sample (relative to a nonabsorbing standard). Most powdered samples must be diluted to the level of a few percent in a nonabsorbing powder to prevent saturation effects on strong absorptions. Figure 3b shows a DRIFTS spectrum of polystyrene powder diluted in KBr after Kubelka-Munk processing. The dilution makes s nearly independent of the sample absorption, so the right-hand side of Equation 12 is usually used as a relative measure of α (18,39,40).

Attenuated total reflection (ATR), also called internal reflection, is based on the phenomenon of total internal reflection. In ATR the infrared beam is directed into an infrared-transmitting crystal so that it strikes the crystal surface at less than the critical angle and undergoes total internal reflection. Often the crystal has parallel sides, so the beam zigzags through the crystal, reflecting repeatedly off the faces. Although the beam is completely reflected at the crystal face, its electromagnetic field extends out beyond the face. This evanescent wave is therefore subject to absorption by a sample in contact with the crystal face. In ATR the sample is placed in intimate contact with the crystal, and the resulting attenuation of the reflected beam is measured. An ATR spectrum is similar to a transmission spectrum and may be treated in the same manner, although the effective path length is imprecisely defined and wavelength dependent. The biggest difficulty in ATR is achieving a reproducible coupling between the crystal and the sample. ATR is therefore best suited for liquids, slurries, pastes, and rubbery or soft solids. Soluble solids can also be analyzed by letting some solution dry on a horizontal crystal face. The depth of penetration, d_p , which is the depth into the sample at which the evanescent wave drops to $1/e$ of its initial magnitude, is given by

$$d_p = \frac{\lambda}{2\pi(n_c^2 \sin^2 \theta - n_s^2)^{1/2}}$$

(13)

where n_c and n_s are the indexes of refraction of the crystal and the sample, respectively, λ is the infrared wavelength (in air), and θ is the angle of incidence. The penetration depth usually falls between 0.05λ and 0.2λ , but n_s increases near absorptions. At a strong absorption, n_s may exceed $n_c \sin \theta$, which means the critical angle is exceeded and the infrared beam refracts out of the crystal instead of reflecting back into it (41–43).

Photoacoustic spectrometry (pas) differs from the other methods in that the detector is a microphone. This makes pas wavelength independent. One detector can be used from the uv through the far ir regions. Pas can be applied both to gases and to nonvolatile condensed-phase samples. In pas the sample is sealed in a small microphone-containing chamber through which an intensity-modulated infrared beam passes. The modulation inherent in ftr spectrometers is well suited for pas. Using a gas sample, the modulated beam passes through and is absorbed by the gas according to its absorption coefficient. This produces a modulated heating and pressurizing of the gas. The microphone picks up the pressure modulations as sound waves. The strength of the signal (the loudness of the sound) is proportional to the strength of the absorption. As of 1994 no free-standing pas accessory designed for gases is commercially available, but instruments containing a gas-sample pas cell and either an ftr spectrometer or a set of bandpass filters are on the market. These have limits of detection for typical organic vapors in air of 10 to 0.1 ppmv. The pas of solid samples is more complex than that of gases. The sample is sealed in a chamber filled with infrared-transparent gas and the modulated beam is aimed onto the sample. Dry, CO₂-free air is an acceptable chamber gas, but helium roughly doubles the signal because of its superior thermal characteristics. Most solid samples are sufficiently opaque to completely absorb the beam. The intensity of the beam falls off exponentially with depth into the sample having a decay constant of $1/\alpha$, where α is the absorptivity (according to Beer's law, $\alpha = ac$). Before the absorbed energy can produce a pas signal, it must diffuse as heat to the sample surface and transfer to the gas. The deposited heat is a damped thermal modulation having a thermal decay coefficient, a_s , given by $a_s = (\pi f D)^{1/2}$, where f is the modulation frequency and D is the thermal diffusivity of the sample. The thermal decay means that the penetration depth for pas is $1/a_s$, because only energy deposited within roughly that distance of the surface can contribute to the signal. As long as $1/a_s \ll 1/\alpha$, the signal is proportional to α , even if the sample is opaque. Typical penetration depths are a few micrometers to a few tens of micrometers. Because the penetration depth for pas varies as $f^{-1/2}$, it can be adjusted over a wider range than the penetration depths of the reflection methods. Figure 3c shows a pas spectrum of polystyrene. Very small (a few tens of micrometers wide) samples may be analyzed by pas without special focusing optics if these are thermally isolated from all but the surrounding gas by mounting them on fine needles. Pas detectors for condensed-phase samples are available commercially (18,44).

In emission spectrometry, the sample is the infrared source. Materials emit infrared radiation by virtue of their temperature. Kirchhoff's law states that the amounts of infrared radiation emitted and absorbed by a body in thermal equilibrium must be equal at each wavelength. A blackbody, which is a body having infinite absorptivity, must therefore produce a smooth emission spectrum that has the maximum possible emission intensity of any body at the same temperature. The emissivity, ϵ , of a sample is the ratio of its emission to that of a blackbody at the same temperature. Infrared-opaque bodies have the same emissivity at all wavelengths so they emit smooth, blackbody-like spectra. On the other hand, any sample dilute or thin enough for transmission spectrometry produces a structured emission spectrum that is analogous to its transmission spectrum because the emissivity is proportional to the absorptivity at each wavelength. The emissivity is calculated from the sample emission spectrum, E_s , by the relation

$$\epsilon = (E_s - E_{bg}) / (E_{bb} - E_{bg})$$

(14)

where E_{bb} is the emission spectrum of a blackbody at the same temperature as the sample and E_{bg} is the background emission observed from the spectrometer and surroundings. An infrared detector is opaque and follows Kirchhoff's law, so it must be cooler than the sample to have a net absorption and produce a signal. The greater the temperature difference between the sample and the detector, the stronger the signal. The emission signal for sample temperatures of a few hundred degrees C or less is much weaker than the signal attained in transmission spectrometry, so transmission is generally preferred. Emission spectrometry is principally used for remote samples, for solids when transmission is not applicable, eg, thin coatings on metal substrates, and for hot gases (18,45).

Infrared microspectroscopy is the union of infrared spectrometry and microscopy for the analysis of microsamples via transmission or reflection (specular, diffuse, or ATR). It requires a specially designed microscope, but the spectrometer need not be specialized, although an instrument having a bright source and high throughput helps offset the losses in the microscope. The microscope must focus the infrared beam from the spectrometer on the sample, collect the infrared radiation that has interacted with the sample, deliver it to a detector, and allow the user to view and define the area to be analyzed. This last need means it must also be a visible-light microscope having identical geometries for the infrared and visible beams. The condenser and objective are mirror-based Cassegrainian optics, so there is no chromatic aberration between the visible and infrared beams. Although the optical geometries of the two beams are identical, the infrared imaging is inferior to the visible because the longer infrared wavelengths are more susceptible to diffraction. The diffraction limit for mid-infrared imaging is roughly 10 μ . Microscopes use one or two apertures to define the analysis area. An aperture before the condenser defines the illuminated area; one after the objective defines the viewed area. Refraction by either the sample or its support can defocus the beam during transmission measurements. Samples for transmission microscopy should generally be 10 to 20 μ thick for adequate but not excessive absorption, so even small samples may need to be flattened or thinned. The type of reflection microscopy used is dictated by sample properties. Smooth samples may be analyzed by specular reflectance and rough surfaced or powdered ones by diffuse reflectance. Intermediate samples produce a combination

of the two reflection types, which must be interpreted with caution. Some instruments permit grazing-angle specular reflectance. Qualitative analysis by infrared microspectrometry is well established, but quantitative analysis is not well developed (46,47).

Remote sensing is the measurement of changes in the open air at a distance. Infrared transmission spectrometry with an open beam path is a common remote-sensing method, and infrared emission can be used if the monitoring target, such as a stack-gas plume is hot. For open-path transmission spectrometry, the elements of a conventional spectrophotometer are rearranged. The infrared source is separated from the rest of the instrument and fitted with a telescope to produce a wide, collimated beam along the monitoring path, which acts as the sample cell. The spectrometer is also fitted with a telescope to collect the beam, which can then be detected and analyzed in the normal manner. Often source-beam modulation and phase-sensitive detection are used to differentiate between the source beam and background infrared sources and improve the signal-to-noise ratio. The simplest arrangement is the bistatic configuration, in which the infrared source and the spectrometer are on opposite ends of the monitoring path, but this requires a power source and very stable mounting at both ends of the path. The monostatic configuration has both the source and the spectrometer at one end of the path and a set of retroreflectors at the other, so the beam traverses the path twice. The monostatic configuration is less vibration sensitive and it simplifies synchronization when phase-sensitive detection is used. Open-path transmission measures the total absorbance along the beam path, so the measurement determines the length-weighted average concentration along the path. Open-path measurements are therefore usually given in units of concentration times length. Typical limits of detection are 1 to 100 ppmv m. A special form of open-path spectrometry is a variant of light detection and ranging (lidar) called differential absorption lidar (dial). In lidar a beam is aimed at the plume and the backscatter resulting from absorption is measured. In dial, the infrared source is normally a laser capable of emission at two nearby wavelengths. One of the wavelengths is within the absorption band of the species to be monitored and the other is outside the band. The difference in absorption between the two wavelengths is largely independent of any broad-band interferences, eg, scattering by dust and background sources. The CO₂ laser is the most common source for infrared dial because it functions in the fingerprint region (48,49).

Open-path measurements are subject to interferences not encountered in the laboratory. Infrared transmission through the open air is subject to absorption and scattering independent of the analysis target. Figure 4 shows the typical transmittance through clean air at sea level. Water and carbon dioxide are responsible for most of the absorptions present. Rayleigh scattering, produced by particles much smaller than a wavelength (principally molecules), scales as the negative fourth power of the wavelength, so it is much weaker in the infrared region than in the visible. On the other hand, Mie scattering, produced by particles roughly the size of a wavelength, can be substantial. Mie scattering is more complicated than Rayleigh scattering because it depends on the relative size of particles and wavelengths (51).

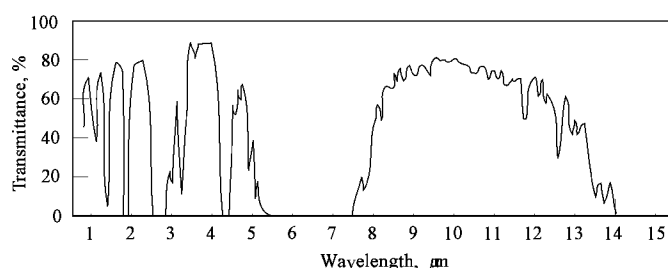


Fig. 4. Transmission of infrared radiation through 1.852 km (1 nautical mi) of air at sea level. Reprinted with permission (50).

Data Analysis. The computerization of spectrometers and the concomitant digitization of spectra have caused an explosive increase in the use of advanced spectrum analysis techniques. Data analysis in infrared spectrometry is a very active research area and software producers are constantly releasing more sophisticated algorithms. Each instrument maker has adopted an independent format for spectrum files, which has created difficulties in transferring data. The Joint Committee on Atomic and Molecular Physical Data has developed a universal format for infrared spectrum files called JCAMP-DX (52). Most instrument makers incorporate in their software a routine for translating their spectrum files to JCAMP-DX format.

Noise and sloped or curving baselines are the most common artifacts. Smoothing is a universally available form of noise removal. Most smoothing algorithms assign to the point being smoothed a weighted average of the data points near it, which reduces the resolution of the spectrum. Different kinds of smoothing, eg, Lorentzian, Gaussian, and Savitsky-Golay, simply involve different weighting functions. Baseline correction takes many forms, but all involve specifying a line or curve which is subtracted from the spectrum. Difference spectra, taking derivatives, Fourier self-deconvolution, and curve fitting are the principal methods for dealing with overlapped bands. Generating a difference spectrum by subtracting one spectrum from another is most commonly used to remove bands of individual components from spectra of mixtures. Difference spectra are also used for depth profiling with sampling techniques having adjustable penetration depths. Most spectrometers have an interactive subtraction routine that allows the user to adjust the relative weighting of the input spectra while viewing the result. Taking an even-order derivative of a spectrum separates overlapping peaks, but taking derivatives lowers the signal-to-noise ratio and introduces side lobes around real peaks, which can hide small peaks. The closer together the peaks are, the higher the order required for separation, but each time the derivative order is increased by two, the signal-to-noise ratio decreases by roughly an order of magnitude. Fourier self-deconvolution narrows spectrum features. It is based on the fact that the Fourier transform of the convolution of a pair of functions is the product of the Fourier transforms of the separate functions. The Fourier transform of a real spectrum of broad peaks, ie, its interferogram, is divided by the Fourier transform of the assumed broad-peak lineshape, and the result is Fourier transformed back into a spectrum (now containing narrowed peaks). It is equivalent, when an ftr spectrometer is used, to re-apodization of the data. Curve fitting is a method of modeling a real absorption band on the assumption that it consists of a series of overlapped peaks having a specific lineshape. Typically the user specifies the number of peaks to attempt to resolve and the type of lineshape. The program then varies the positions, sizes, and widths of the peaks to minimize the difference between the model and the spectrum. The largest difficulty is in knowing the correct number of peaks to resolve. Derivative spectra are often useful in determining the correct number (18,53,54).

Qualitative analysis is the oldest application of infrared spectrometry. Unknowns are identified from their infrared spectra by the use of either group frequency correlations or spectral libraries. Spectral libraries are available both in book form (55) and as computer databases (qv). Both spectrometer makers and third party vendors supply spectral databases and software for searching them. At present, the largest commercial database, Sadtler Condensed Phase IR Standards, contains 69,000 spectra. Many databases and software can be used with personal computers, but on-line databases exist as well. The search software can either provide the spectrum of a user-designated compound, or produce a list of the closest matches (hits) to the spectrum of an unknown. A single spectrum represents a large amount of data, thus the simplest libraries and search routines use tables of band positions and strengths, rather than full spectra. Advanced search algorithms often use such a reduced data set for the first pass so as to speed searching. Many libraries also contain physical data, such as boiling point, melting point, and density, which may be searched in conjunction with the spectra. Some of the most advanced search software is multispectral; that is, it can search multiple libraries containing complementary data, such as mass spectra and nmr spectra. Most libraries contain transmission or absorption spectra. Although spectra acquired by other sampling methods can be converted to transmission or absorption form, these conversions are often imperfect, so spectra acquired by different sampling methods may have peaks of different relative intensities. Matching unknown spectra taken by one sampling method to library entries generated by another may best be done using only peak locations. Library searching is a very powerful method for the analysis of good quality spectra of relatively pure compounds or standard mixtures. Library searching is not so successful using spectra either of unknown mixtures or containing artifacts (56,57). Books on group frequencies are available for manual identifications (24,26), but software containing correlation data has been developed for automatic or computer-assisted identification. Group frequency correlations never match all peaks in a spectrum, because not all peaks arise from specific groups. Accordingly, group frequency correlations are more tolerant of mixtures and artifacts.

The result of a group frequency identification can be used as a prefilter for library searches (57–59).

The classic method of quantifying using Beer's law (eq. 10) is still the preferred method where applicable. A number of multivariate analysis methods have been developed to make use of data from many wavenumbers (even the complete spectrum) when a single wavenumber is inadequate. Most of the methods use a set of calibration spectra (a training set) to generate a model of how the spectral data are related to parameters of the calibration set. The training set must span all sample variability the model later encounters while analyzing unknowns; multivariate techniques are excellent at interpolating between standards, but they are poor at extrapolating beyond the limits of the training set. Once the model is generated, an unknown may be analyzed by applying the model to its spectrum. The most common of the multivariate techniques are also factor analysis methods, which are matrix algebra methods in which the starting data set is considered the product of a data set of factors (scores) and a matrix of factor loadings (loading vectors) (see CHEMOMETRICS) (18,60–62).

Hyphenated Techniques. Hyphenated techniques are methods in which two or more analytical technologies have been joined together in a single process (see ANALYTICAL METHODS, HYPHENATED INSTRUMENTS). Most commonly, the combination is a separation technique with an instrumental analysis technique, eg, infrared spectrometry, used as a detector. Development of hyphenated infrared techniques has centered on combining a chromatographic method with ftr (see CHROMATOGRAPHY). Commercial instruments are available that combine ftr with gas chromatography (gc ftr), high performance liquid chromatography (hplc ftr), and supercritical fluid chromatography (sfc ftr). These combine the power of chromatography to handle mixtures with the identification ability of infrared analysis. Gc ftr is usually accomplished using a light pipe, typically a thin, heated, gold-coated tube with windows at each end, attached to the end of the column. Gc ftr instruments can record full mid-infrared or near infrared spectra in real time. Matrix isolation has also been used for a gc ftr interface. The gc eluent and argon are codeposited on a chilled, slowly rotating, reflective or infrared-transmitting disk. The infrared beam is then focused on the eluent trail to record spectra from the sample spots. This does not produce real-time data, but it can give excellent results, because the chromatographic record is literally frozen in place. Any desired sample spot may be positioned in the infrared beam, and a large number of scans can then be co-added to achieve high sensitivity. Commercial instruments that add mass spectrometry to the combination, gc ftr ms, are also available. The column effluent is split between the two spectrometers and their complementary analyses can then be used in tandem. Gc ms is generally good for differentiating homologous compounds but poor for isomers, whereas gc ftr is good for isomers but poor for homologues. Interfacing hplc or sfc with ftr is more complex because the mobile phase is usually not infrared transparent. Nevertheless, flow cells analogous to ordinary liquid-sample cells are available for hplc ftr and sfc ftr and give good results with the proper mobile phases. ATR cells can also be used as interfaces. These reduce solvent contributions to spectra. Several methods involving elimination of the mobile phase have been demonstrated for cases in which the mobile phase is considerably more volatile than the analytes. Most are analogous to the matrix-isolation interface used for gc ftr (without the argon) and involve spraying the eluent onto a moving reflective or transparent surface held at a temperature that allows the mobile phase to evaporate. Chromatography using a traditional detector produces a chromatogram, the time plot of a single variable. Infrared analyzers provide many variables, thus many different chromatograms can be constructed. Absorbances at selected wavelengths are obvious choices. The Gram-Schmidt method of interferogram-based chromatogram construction is popular because it is fast, enhances signal-to-noise ratios, and need not be limited to certain analytes (63–65). Thermogravimetric analysis (tga) is a nonchromatographic method which readily combines with ftr. In tga a weighed sample is slowly heated and its weight loss as a function of temperature is recorded. An ftr spectrometer can be interfaced to record spectra of the gases evolved during the pyrolysis so that the products and reactions may be determined.

Applications. The most ubiquitous use of infrared spectrometry is chemical identification. It has long been an important tool for studying newly synthesized compounds in the research lab, but industrial identification uses cover an even wider range. In many industries ir spectrometry is used to assay feedstocks (qv). In the flavors (see FLAVORS AND SPICES), fragrances (see PERFUMES), and cosmetics (qv) industries, it can be used not only for gross identification of feedstocks, but for determining specific sources. The spectra of essential oils (see OILS, ESSENTIAL), essences, and other natural products vary with the season and source. Adulteration and dilution can also be identified.

The polymer industry uses infrared spectrometry for both quality control and research. The chemical effects of surface coatings (qv) can be isolated from the spectrum of the treated material by subtracting out the spectra of the untreated substrate and the pure coating material. The blend of two incompatible polymers is a two-phase system; its spectrum is simply the sum of the separate homopolymer spectra. The blend of compatible polymers produces a chemical interaction between the polymers. The spectrum of this interaction can be isolated by subtracting out the homopolymer spectra. More intimate mixing, as in block and random copolymers, produces more changes in the spectra. Degradation and weathering can be studied by the spectrum changes these cause. Faster processes can also be studied because modern spectrometers can record 30 to 50 complete spectra (or interferograms) per second. Instruments built around multichannel analyzers can operate even faster. Spectra of chemical reactions, such as uv-curing, and physical changes, such as the application of strain, can be recorded in real time (18,66).

Microspectroscopy applies the identification power of infrared spectroscopy to the microscopic realm. Contaminants on printed circuit boards, blemishes in coatings, and other production defects can be isolated *in situ* and analyzed (see ELECTRONICS, COATINGS). Analysis of flaws that develop during use illuminates the method of failure. Microscopic samples, such as particulates filtered from air, can be analyzed individually. The forensic applications are many: paint chips, single fibers, explosive residues, and inks on currency can all be identified nondestructively (see FORENSIC CHEMISTRY). The structures of layered materials, such as laminated polymer films, are studied via microspectroscopy by cross-sectioning the materials and examining the individual layers edge on (47).

Gc ftr has both industrial and environmental applications. The flavor and aroma components in fragrances, flavorings, and foodstuffs can be identified and quantified via gc ftr (see FOOD ADDITIVES). Volatile contaminants in air, water, and soil can be analyzed. Those in air are usually trapped in a sorption tube then injected into the chromatograph. Those in water or soil are sparged, extracted, or thermally desorbed, then trapped and injected (63,64).

Process monitoring, both in the laboratory and on the production line, is an important application of infrared spectrometry (see PROCESS CONTROL). In the laboratory a reaction can be monitored *in situ* to determine how it proceeds or to optimize its performance before scale up. Sampling accessories built into reaction vessels or designed for insertion into vessels are commercially available. On-line infrared monitoring dates from the 1950s, when nondispersive instruments, ie, filter spectrometers having a single bandpass filter, were first used. Modern on-line devices are designed for the plant, with automatic sampling, optical benches isolated from the environment, and software structured for continuous rather than batch operation. The optical bench and sampling interface can be physically separated through the use of optical fibers or light pipes. Complete ftr spectrometers for on-line use are commercially available, but filter-wheel devices are the most common because of simplicity, compactness, and ruggedness. A wide variety of on-line sampling interfaces are commercially available, including dip and insertion probes, based on transmission, diffuse reflection, or ATR; flow cells for gases, liquids, and polymer melts; diffuse reflection devices for solid process streams; and sparging systems for wastewater streams. The properties that can be monitored cover a wide range: additive content, blending, and crystallinity in polymers, process gas composition, octane number of gasolines (see GASOLINE AND OTHER MOTOR FUELS), cure level of coatings and thermosets, and identification of reaction endpoints (67).

Near infrared spectrometry has been particularly popular in the feed and food processing (qv) industries where its large depths of penetration allow analysis of processes *in situ* and of intact or minimally processed materials (see FEEDS AND FEED ADDITIVES). Its use is gaining popularity in other industries. Reflection techniques are most common, but transmission is also used. It is used to determine moisture, fat, and protein levels in food products ranging from dry milk, to meat, to bread. It is used to determine protein, moisture, and bran separation in grab samples during grain milling to identify the processing endpoint. It has been used to determine starch granule damage, loaf score, and other physical and chemical properties of the resulting flours. Unsaturation levels in edible oils are monitored with near infrared spectrometry. It can monitor fermentation (qv) processes; alcohol, protein, total nitrogen, ash, and many other properties can be determined *in situ* with a probe or through a window (33,34). In the pharmaceutical industry, moisture levels during granulation and homogeneity during mixing are monitored so as to determine the endpoints of these processes. It can also assay intact tablets. Degradation products, residual moisture, and even dissolution rates for finished products have been determined using near infrared analysis (68). The long-path lengths available in the near infrared region allow its noninvasive biomedical use. Near infrared transmission through a finger has been used to analyze blood chemistry, including hemoglobin level, hematocrit, blood glucose, blood urea nitrogen, and bilirubin.

The environmental applications of infrared spectrometry are many and varied. Many applications at industrial sites are analogous to those for on-line process analysis; waste streams and recycling processes can be monitored in the same way. Commercial infrared stack-gas monitors are based on either an extractive probe attached to a long-path gas cell or an open-path (across stack) configuration (69). Stack plume and flare monitoring can be done externally

using open-path emission spectrometry (49). Both point-sampling devices and open-path spectrometers are also used for general emission monitoring. An open-path system can be set up along the perimeter of a facility or an environmental remediation operation to monitor all emissions leaving the site (48). Point-sampling devices range from simple hand-operated single-filter nondispersive devices to self-contained automatic-sampling fir spectrometers. Industrial hygiene (qv) applications of point-samplers are numerous. Wherever gases or volatiles may be present, point samplers can be used to monitor worker exposure and to comply with OSHA or other regulations (70). On a global scale, infrared remote sensing is done from satellites.

Radiometry

The nomenclature in radiometry is not completely standardized. Often several terms are used for the same thing. The radiant flux is the flow of energy in watts. Radiant exitance and radiant emittance both mean the radiant flux per unit area emitted from a surface. On the other hand, irradiance and radiant incidence both mean the radiant flux per unit area incident on a surface. Radiant areance also means the radiant flux per unit area, but its use is not limited to a specific direction. Radiant intensity and radiant pointance mean the radiant flux per unit solid angle emitted from a surface. Radiance and radiant stearence mean the radiant flux per unit area per unit solid angle passing through any surface normal to the direction of the flux. Replacing the word *radiant* in the above terms by *photon*, produces the analogous terms for energy flow in photons instead of watts. The above terms are usually used to mean spectrally integrated quantities, although the adjective *total* may be added to mean this explicitly. Similarly, the adjective *spectral* may be added to define a term as a function of wavelength or wavenumber, determined per unit wavelength or wavenumber. A more complete nomenclature is given in the literature (71–73).

Thermal Emission Laws. All bodies emit infrared radiation by virtue of their temperature. The total amount of radiation is governed by Kirchhoff's law, which states that a body at thermal equilibrium, ie, at the same temperature as its surroundings, must emit as much radiation as it absorbs at each wavelength. An absolutely blackbody, one that absorbs all radiation striking it, must therefore emit the most radiation possible for a body at a given temperature. The emission of this so-called blackbody is used as the standard against which all emission measurements are compared. The total radiant emittance, *M*, for a blackbody at temperature *T* is given by the Stefan-Boltzmann law,

$$M = \sigma T^4$$

(15)

where $\sigma = 5.670 \times 10^{-12} \text{ W (cm}^2\text{K}^4\text{)}$. The spectral radiant emittance is given by Planck's law:

$$M_\lambda = \frac{2\pi hc^2}{\lambda^5 (e^{hc/\lambda kT} - 1)} \quad M_{\tilde{\nu}} = \frac{2\pi h c^2 \tilde{\nu}^3}{e^{hc\tilde{\nu}/kT} - 1}$$

(16)

where *M_λ* and *M_{ν~}* are the spectral radiant emittance per unit wavelength and per unit wavenumber, respectively, *λ* and *ν~* are the wavelength and wavenumber of the radiation, respectively, *h* is Planck's constant, *k* is Boltzmann's constant, and *c* is the speed of light. Figure 5 shows plots of these two versions of Planck's law. The emittance increases at all points and the peak moves to higher energy with increasing temperature. The wavelength or wavenumber of the peak is given by Wien's Displacement law:

$$\lambda_{peak} T = 0.2898 \text{ cm} \cdot \text{K} \quad \frac{\tilde{\nu}_{peak}}{T} = 1.961 \text{ (cm} \cdot \text{K)}^{-1}$$

(17)

All bodies other than blackbodies reflect or transmit some portion of the radiation incident upon them. The fractions of the incident radiation that are absorbed, transmitted, and reflected by a body are called the absorptivity, *α*, the reflectivity, *ρ*, and the transmissivity, *τ*, respectively. These must add up to one:

$$\alpha + \rho + \tau = 1$$

(18)

The emissivity, *ε*, is the ratio of the radiant emittance of a body to that of a blackbody at the same temperature. Kirchhoff's law requires that *α = ε* for all bodies at thermal equilibrium. For a blackbody, *α = ε = 1*. Near room temperature, most clean metals have emissivities below 0.1, and most nonmetals have emissivities above 0.9. This description is of the spectrally integrated (or total) absorptivity, reflectivity, transmissivity, and emissivity. These terms can also be defined as spectral properties, functions of wavelength or wavenumber, and the relations hold for the spectral properties as well (71,74–76).

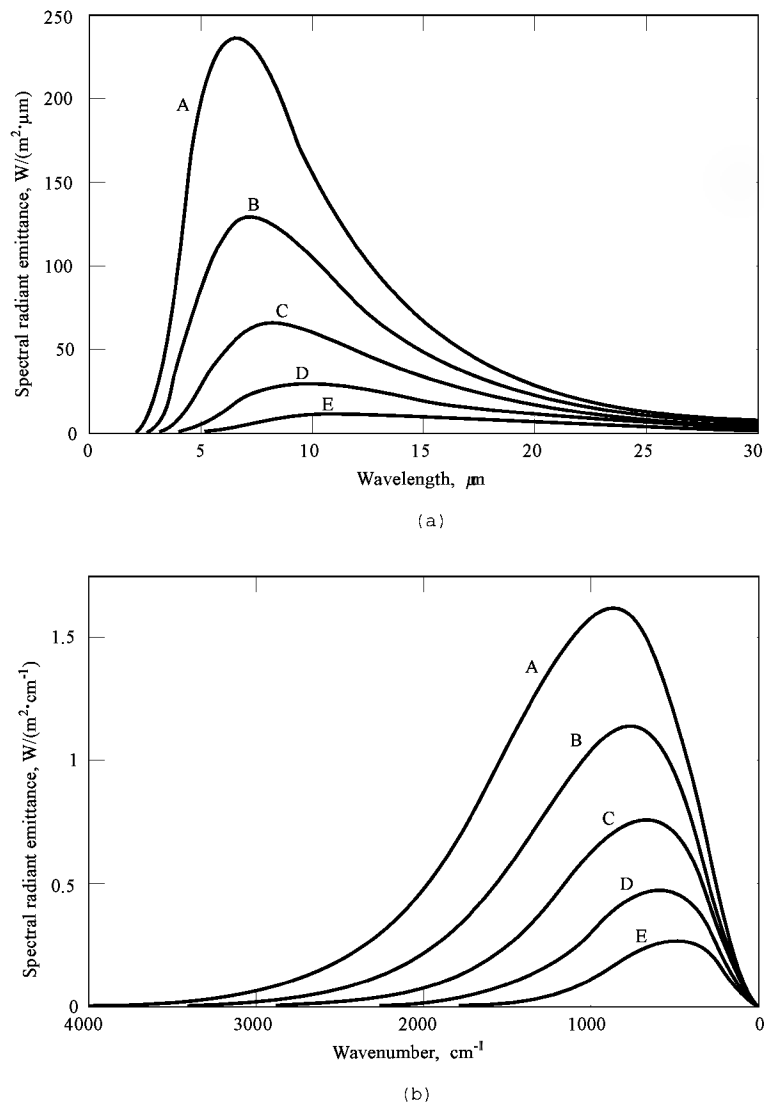


Fig. 5. Spectral radiant emittance of a blackbody as predicted by Planck's law for A, 450 K; B, 400 K; C, 350 K; D, 300 K; and E, 250 K per unit (a) wavelength and (b) wavenumber.

Radiometric Measurements. The basic radiometric parameter is the radiance, because all other quantities can be determined from it by integration. Radiance has the special property that it is constant along any line radiating out from a source, excluding absorption and scattering losses. That is, the radiance at a point on the surface of a source equals the radiance at any point on a line extending out from that point measured along the direction of that line. What a radiometer directly measures is the radiant flux, ϕ , striking its detector, but this can be directly related to source radiance because of its property of constancy. The simplest possible radiometer is a circular tube of length l and cross-sectional area $\mathcal{A}$ with one open end and a circular detector of area a mounted at the other end. The open end of the tube therefore defines the limiting aperture of the radiometer, and the solid angle, Ω , of the detector viewed from this aperture is a/l^2 . First, consider the case where the radiometer is observing a flat source normal to its surface, the source is large enough to fill the field of view of the radiometer, and it has uniform radiance, L . The radiance at every point of the limiting aperture is also L because of the constancy of the radiance. The flux reaching the detector from a single point in the aperture is therefore $L\Omega$, and the total flux is given by $\phi = LA\Omega$. As long as the source completely fills the radiometer field of view, the measured flux is independent of the separation between the source and the radiometer. Now consider the case where the source of area A_s and range r from the radiometer does not fill the radiometer field of view, but all other conditions are as before. The radiance at every point in the radiometer aperture is still L , but the flux reaching the detector from a single point in the aperture is now $L\Omega_s$, where $\Omega_s = A_s/r^2$, the solid angle of the source viewed from the limiting aperture. This means $\phi = LA\Omega_s = LA A_s/r^2$. This is the principal relation between radiance and measured flux. If the source is not flat or is not observed normal to its surface, then A_s is the area of the source projected onto a plane normal to the direction of observation. If the source radiance is not uniform, the radiometer measures the area-weighted average of L (72,77).

The radiance observed from a body depends on the elevation angle θ , which is the angle from the surface normal at which the measurement is made. If ϵ is independent of θ , the radiance at θ , is given by Lambert's cosine law:

$$R_{\theta} = R_n \cos \theta \tag{10}$$

where R_n is the radiance observed at the surface normal. Although Lambert's law is a useful approximation, the emissivity does depend on θ for most surfaces. For many dielectric surfaces, the emissivity at θ , ϵ_{θ} , obeys the relation

$$\epsilon_{\theta} = [4n(1 + n)^2] \cos^a \theta \tag{20}$$

where n is the index of refraction and a is an empirically determined constant less than one (78).

For the ideal case discussed, there are no interferences from the surroundings, the atmosphere has no effect, and all radiation from the target results from its emission. In the real world, the radiometer can observe emission not only from the target, but from other emission sources, including the radiometer itself. Most radiometers can compensate for emission from themselves. To eliminate direct contributions from surroundings, the image of the target must overfill the detector so that blurring effects do not result in measurement errors. If the target has an emissivity less than one, it can reflect radiation from its surroundings. The lower the emissivity, the greater the reflectivity. Some radiometers can compensate for this if an emissivity value for the target and the ambient temperature are provided. Alternatively, the radiance from a highly reflective test target can be measured to determine the contribution from surroundings. The intervening atmosphere absorbs, emits, and scatters infrared radiation. These effects depend on humidity and haze, but the closer the target is, the less effect the atmosphere has. Some instruments have the ability to compensate for these interferences if ambient temperature, humidity, and range to the target are known (79–82).

The wavelength range used can have a large effect on a radiometric measurement. Incorporating a filter in the radiometer can tailor it for a specific measurement task. For example, a natural gas flame is nearly transparent at 3.7 μm, so at this wavelength the walls of a furnace may be measured through flame. In humid environments, a wavelength band less affected by water absorption can be chosen (see Fig. 4). Most radiometric and thermographic measurements are made either in the mwir (3–8 μm) or in the lwir (8–13 μm). The latter has a number of advantages, including greater sensitivity for targets near room temperature (see eq. 17). A 290 K target emits 127 W m² in the lwir but only 4.1 W m² in the mwir. By contrast, solar radiation on a clear day is 16 times greater in the mwir than the lwir, so the lwir is less susceptible to reflected radiation during outdoor measurements. The lwir is also more sensitive to target temperature changes because the temperature derivative of the spectral emittance (dM_{λ}/dT) is greatest in the lwir for temperatures in the vicinity of room temperature. The lwir is scattered less by mist and smoke, and is less affected by humidity because it is absorbed less by water vapor. The mwir also has its advantages. The thermal contrast ($[dM_{\lambda}/dT]/M_{\lambda}$) is greater in the mwir. Measurements in the mwir are also less susceptible to the temperature of the surroundings. The mwir is more sensitive than the lwir for targets at high temperatures. The mwir provides better image resolution because it is less susceptible to diffraction and larger imaging arrays are available, mwir detectors generally have higher D^* s than lwir detectors, so warmer detectors can be used in the mwir (15,79,80).

Applications. Applications of radiometers and thermal imagers range from energy audits of buildings, to airborne crop surveys, to helping firefighters locate people in smoke-filled rooms. The applications related to chemical technology primarily deal with process monitoring and plant maintenance. Historically, many of these applications have been more fully utilized in other areas, particularly steel (qv) and electronics, but as engineered materials continue to replace metals and as quality control needs become more stringent, they are increasingly being adopted in chemical activities. The most obvious plant inspection tasks that radiometers and thermal imagers can be used for involve elevated or reduced temperature processes (see TEMPERATURE MEASUREMENTS). A thermal imager simplifies inspection because large areas can be surveyed rapidly and the image is normally much simpler to interpret than multiple point measurements. Abnormal temperature distributions in heat exchangers or air coolers can indicate blocked or constricted pipes. Leakage by a steam trap produces a local hot spot. Thermal inspection of furnaces can reveal poor flame patterns, inefficient combustion, and coking. The outer shells of insulated vessels inspected during operation show hot or cold spots wherever flaws occur in the refractory or insulating linings. The temperature determined at such a spot is a measure of the threat to the outer shell. A second group of applications involve detecting the heat produced by malfunctions or incipient failures. Worn bearings produce hot regions in mechanical equipment. Poor contacts in electrolytic cells are observable. Stressed transformers, breakers, and other electrical gear radiate at the stress points (81).

Many plant applications rely on the fact that fluids in pipes and vessels are rarely at ambient temperature, even if not part of an elevated or reduced temperature process, so their presence (or absence) can be detected thermally. Reduced fluid flow in a pipe alters the surface temperature of the pipe; it is closer to ambient. A localized deposit within a pipe produces an area at a temperature closer to ambient because of the insulating effect of the deposit. The level to which tanks are filled can be determined because of the distinct temperature change in the tank wall at the liquid level. Even leaks from shallow underground pipes produce observable changes in the thermal pattern of the ground. A gas leak into the open can be detected using a radiometer or thermal imager if the gas absorbs infrared radiation within the spectral response of the instrument and the gas temperature differs from that of the background. Oil spills in water can be detected in the infrared region because oil and water have different emissivities and reflectivities. Normal process operations can also be monitored radiometrically. A thermal imager can be used to image a process web, or a radiometer combined with a line scanner can monitor the full width of a process web. Motion of the web then allows an image to be built up line by line. In the paper (qv) industry, web temperature correlates with moisture level in the paper during the drying process, so the degree and evenness of drying can be monitored. Temperature distributions in molds and dies can be assessed (81).

Radiometry and thermal imaging can also be used in the stress analysis of objects. Radiometric stress analysis has usually been used for metals, but the methods are being adapted for plastics and composites. Compression, tension, or any other stress on an object that produces a volume change produces heat at the stress point because work, in the thermodynamic sense, is done on the object. Pure shear, which produces no volume change, does not cause heating. An object can be tested by applying a cyclic stress and monitoring the cyclic temperature changes induced (83).

Nondestructive evaluation (nde) can be done using radiometry and thermal imaging (see NONDESTRUCTIVE EVALUATION). A thermal gradient is induced in an object either by applying or removing a heat source at one end. The surface temperature distribution is then monitored as the object heats or cools. Flaws and structures within the object affect the observed temperature distribution. Voids, cracks, and delaminations cast thermal shadows. A surface region heats more slowly if such a defect lies between it and the heat source. Ribs, honeycombs, and other internal structures can be visualized because they act as internal heat conduction paths. The disbond of a rib is evidenced by its absence in the thermal image (84). Thermal wave imaging is an advanced form of radiometric nde. It is related to photoacoustic spectrometry, which also depends on thermal waves. A thermal wave is the periodic movement of heat caused by periodic heating. Any boundary, flaw, or structure within the body of an object causes a shift in the phase of the thermal wave. Thermal wave imaging can provide more detail than traditional radiometric nde. For example, coating thickness can be measured from the phase of the thermal wave reflected off the coating-substrate interface. Thermal waves are heavily damped, however, so their depth of penetration is limited and they can only be used for near-surface structures (85).

BIBLIOGRAPHY

"Infrared Technology" in *ECT* 3rd ed., Vol. 13, pp. 337–355, by C. Warren, AGA Infrared Systems AB.

1. G. W. Chantry, *Long-Wave Optics*, Vols. 1 and 2, Academic Press, Inc., Orlando, Fla., 1984, pp. 542–545.
2. J. Hecht, *Laser Focus World* **28**(5), 127–143 (1992); **28**(8), 79–89 (1992).
3. Ref. 1, pp. 600–602.
4. Ref. 1, pp. 594–600.
5. Ref. 1, pp. 591–594.
6. Ref. 1, pp. 547–584.
7. W. L. Wolfe, in W. L. Wolfe and G. J. Zissis, eds., *Infrared Handbook*, rev. ed., Environmental Research Institute of Michigan, Ann Arbor, Mich., 1985, pp. 7-1 to 7-137.
8. P. Klocek and G. H. Sigel, Jr., *Infrared Fiber Optics*, SPIE Optical Engineering Press, Bellingham, Wash., 1989.
9. Ref. 8, pp. 33–36.
10. P. R. Griffiths and J. A. de Haseth, *Fourier Transform Infrared Spectrometry*, John Wiley & Sons, Inc., New York, 1986, pp. 248–254.
11. P. W. Kruse, L. D. McGlauchlin, and R. B. McQuistan, *Elements of Infrared Technology*, John Wiley & Sons, Inc., New York, 1962, pp. 268–277.
12. Ref. 1, Vol. 1, pp. 358–364.
13. Ref. 10, pp. 209–212 and 216–217.
14. Product literature, Cincinnati Electronics, EG&G Judson, Graseby Infrared, and Hamamatsu Photonics, 1993.
15. A. P. Davis and A. H. Lettington, in S. G. Burnay, T. L. Williams, and C. H. Jones, eds., *Applications of Thermal Imaging*, Adam Hilger, Bristol, U.K., 1988, pp. 1–34.
16. Ref. 11, pp. 386–413.
17. I. J. Spiro and M. Schlessinger, *Infrared Technology Fundamentals*, Marcel Dekker, Inc., New York, 1989, pp. 177–183.
18. J. L. Koenig, *Advances in Polymer Science*, Vol. 54, Springer-Verlag, Berlin, 1983, pp. 87–161.
19. Ref. 10, pp. 1–9.
20. Ref. 10, pp. 6–25 and 56–64.
21. Ref. 10, pp. 274–276.
22. C. D. Tran, *Anal. Chem.* **64**, 971A–981A (1992).

23. J. I. Steinfeld, *Molecules and Radiation: An Introduction to Modern Molecular Spectroscopy*, 2nd ed., MIT Press, Cambridge, Mass., 1978.

24. N. B. Colthup, L. H. Daly, and S. E. Wiberley, *Introduction to Infrared and Raman Spectroscopy*, 3rd ed., Academic Press, Inc., San Diego, Calif., 1990.

25. Ref. 24, pp. 109–169.

26. D. Lin-Vien, N. B. Colthup, W. G. Fateley, and J. G. Grasselli, *Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules*, Academic Press, Inc., San Diego, Calif., 1991.

27. A. L. Smith, *Applied Infrared Spectroscopy: Fundamentals, Techniques, and Analytical Problem-Solving*, John Wiley & Sons, Inc., New York, 1979, pp. 140–186.

28. Ref. 27, pp. 137–138.

29. Ref. 27, pp. 161–169.

30. Ref. 24, pp. 100–101.

31. Ref. 27, pp. 219–222.

32. Ref. 24, p. 397.

33. E. Stark, K. Luchter, and M. Margoshes, *Appl. Spec. Rev.* **22**, 335–399 (1986).

34. C. S. Creaser and A. M. C. Davies, eds., *Analytical Applications of Spectroscopy*, Royal Society of Chemistry, London, 1988.

35. Ref. 24, pp. 85–90.

36. Ref. 27, pp. 73–83.

37. Ref. 1, Vol. 1, pp. 214–221.

38. Ref. 10, pp. 187–191.

39. P. J. Brimmer, P. R. Griffiths, and N. J. Harrick, *Appl. Spec.* **40**, 258–265 (1986).

40. Ref. 24, pp. 90–91.

41. Ref. 10, pp. 191–194.

42. Ref. 24, pp. 92–92.

43. Ref. 27, pp. 84–95.

44. J. F. McClelland, R. W. Jones, S. Luo, and L. M. Seaverson, in P. Coleman, ed., *Practical Sampling Techniques for Infrared Analysis*, CRC Press, Boca Raton, Fla., 1993, pp. 107–144.

45. Ref. 10, pp. 202–208.

46. J. E. Katon and A. J. Sommer, *Anal. Chem.* **64**, 931A–940A (1992).

47. K. Krishnan and S. L. Hill, in J. R. Ferraro and K. Krishnan, eds., *Practical Fourier Transform Infrared Spectroscopy*, Academic Press, Inc., San Diego, Calif., 1990, pp. 103–165.

48. R. L. Spellicy, W. L. Crow, J. A. Draves, W. F. Buchholtz, and W. F. Herget, *Spectroscopy* **6**(9), 24–34 (1991).

49. R. L. Spellicy and M. J. Persky, in J. Wormhoudt, ed., *Infrared Methods for Gaseous Measurements*, Marcel Dekker, Inc., New York, 1985, pp. 139–195.

50. Ref. 11, p. 164.

51. Ref. 11, pp. 162–193.

52. R. S. McDonald and P. A. Wilks, Jr., *Appl. Spec.* **42**, 151–162 (1988).

53. P. R. Griffiths, J. A. Pierce, and G. Hongjin, in H. L. C. Meuzelaar and T. L. Isenhour, eds., *Computer-Enhanced Analytical Spectroscopy*, Plenum Press, New York, 1987, pp. 29–54.

54. Ref. 10, pp. 101–106 and 231–232.

55. C. J. Pouchert, *Aldrich Library of FT-IR Spectra*, Vols., 1–3, Aldrich Chemical, Milwaukee, Wis., 1985 and 1989.

56. S. R. Heller and S. R. Lowry, in Ref. 53, pp. 223–237.

57. J. T. Clerc, in Ref. 53, pp. 145–162.

58. A. Savitzky, in Ref. 53, pp. 183–199.

59. H. B. Woodruff, in Ref. 53, pp. 201–222.

60. D. M. Haaland, in Ref. 47, pp. 395–468.

61. W. Windig and H. L. C. Meuzelaar, in Ref. 53, pp. 67–102.

62. Ref. 10, pp. 355–366.

63. R. White, *Chromatography/Fourier Transform Infrared Spectroscopy and Its Applications*, Marcel Dekker, Inc., New York, 1990.

64. R. Namba, in Ref. 47, pp. 469–518.

65. Ref. 10, pp. 564–647.

66. Ref. 27, pp. 186–194.

67. J. Coates, T. Davidson, and L. McDermott, *Spectroscopy* **7**(9), 40–49 (1992).

68. E. W. Ciurczak and J. K. Drennen, *Spectroscopy* **7**(6), 12–13 (1992).

69. J. T. Cronin, *Spectroscopy* **7**(5), 33–39 (1992).

70. K. A. Bunding Lee and co-workers, *Spectroscopy* **8**(5), 24–29 (1993).

71. W. L. Wolfe, in Ref. 7, pp. 1-1 to 1-41.

72. F. E. Nicodemus, *Am. J. Phys.* **31**, 368–377 (1963).

73. Ref. 17, pp. 20–25.

74. A. J. LaRocca, in Ref. 7, pp. 2-1 to 2-97.

75. Ref. 11, pp. 13–37.

76. Ref. 17, pp. 6–19.

77. Ref. 17, pp. 33–52.

78. Ref. 11, pp. 42–48.

79. J. T. Woolaway, *Photonics Spectra* **25**(2), 113–119 (1991).

80. B. Linnander, *Laser Focus World* **29**(6), 95–99 (1993).

81. T. L. Williams, in Ref. 15, pp. 35–72.

82. Ref. 17, pp. 20–21.

83. J. M. B. Webber, in Ref. 15, pp. 191–225.

84. R. Paulson, A. Aquino, H. Decker, and T. Schapp, in I. J. Shapiro, ed., *Modern Utilization of Infrared Technology VIII*, SPIE Proceedings, Vol. 366, SPIE, Bellingham, Wash., 1983, pp. 88–95.

85. D. P. Almond and P. M. Patel, in Ref. 15, pp. 226–242.

Roger W. Jones
Ames Laboratory, USDOE

RAMAN SPECTROSCOPY

Photons can interact weakly with a molecule, resulting in the excitation of a molecular vibration, and the subsequent scattering of a new photon at a slightly longer wavelength (lower energy) than the incident photon. This effect is called Raman scattering, and its measurement, Raman spectroscopy. The exciting light is usually a visible or near-infrared (nir) laser (see LASERS). Raman scattering can also occur upon excitation of molecular rotations or electronic transitions; however, only vibrational Raman spectroscopy has found widespread practical application. Thus the term Raman spectroscopy is used herein to mean vibrational Raman spectroscopy.

Although it had been predicted several years earlier, the first observation of Raman scattering was published in 1928 by C. V. Raman and K. S. Krishnan (1). This inelastic scattering of visible light, an effect analogous to the Compton effect for x-rays, was initially called feeble fluorescence, although fluorescence involves the reradiation of absorbed light and in Raman scattering no absorption takes place. In 1930, Raman was awarded the Nobel Prize in Chemistry and this inelastic scattering effect was given his name. Since then, Raman spectroscopy has been applied to many areas of industrial, medical, and academic interest.

Although a handful of industrial laboratories have employed Raman spectroscopy for many years, interest has blossomed since the early 1980s. Both the solution of several vexing experimental problems and the development of sophisticated user-friendly instrumentation have contributed to this growth. Specialist reviews of Raman work in most primary areas of application exist, including reviews for inorganic materials (2,3), semiconductors (qv) and superconductors (4,5), polymers (6,7), process control (qv) (8), remote sensing (usually through a fiber-optic probe (see FIBER OPTICS)) (9), and biological systems (10–12). Moreover, comprehensive reviews of the most recently published material in the field appear biennially (13,14).

The first Raman experiments used sunlight as the excitation source, telescope lenses to collect the scattered light, and the human eye to detect the weak, Raman-shifted light. Shortly afterward, arc lamps and high power mercury lamps were used with the appropriate filters to generate a monochromatic beam, and the scattered light was recorded on photographic plates. These experiments took anywhere from several hours to more than a day to complete. The introduction of lasers in the early 1960s and the advent of photomultiplier tubes and chart recorders reduced measurement time to 10–30 min. As of the 1990s using compact lasers, which provide intense monochromatic light, efficient spectrographs, and sensitive charge-coupled device (CCD) detectors under computer control, Raman spectra can sometimes be recorded in times as short as a few ms and are routinely recorded in a few seconds.

Theory

Every molecule generates its own characteristic vibrational Raman spectrum, which can be used for qualitative identification. Raman scattering arises from interaction between the electromagnetic field of a photon and the electric field of the electron cloud of a molecule. The strength of the scattering depends on the magnitude of the change in deformation of the electron cloud in the external field, caused by a molecular vibration. The magnitude is given by the (spatial) derivative of the molecular polarizability. The fundamental selection rule in Raman spectroscopy is that the vibration must cause a change in the polarizability of the molecule. This rule differs from the infrared absorption selection rule, which is that the vibration must cause a change in the dipole moment of the molecule. This difference in selection rules leads to a difference in the appearance of Raman and infrared spectra. For example, totally symmetric vibrations of centrosymmetric molecules are Raman-active but infrared-silent. Examples include the C–C bond stretching of carbon tetrachloride and the C–C bond stretching of ethylene. Asymmetric vibration of centrosymmetric molecules are Raman-silent.

For any molecule where structure and hence symmetry point group are known, the application of the selection rules allows the prediction of the number of fundamental Raman- and infrared-active vibrational modes, as well as which modes can be observed in both spectra. Generally, vibrations of highly polar moieties, such as the hydroxyl group, are weak in the Raman and strong in the infrared. C–H bond stretching and C–C stretching and aromatic ring breathing are frequently strongly Raman active. Carbonyl stretches and C–C stretches have moderate Raman intensity.

It is experimentally easy to generate Raman spectra using polarized light and to observe the partial depolarization of the spectra. Bands of totally symmetric vibrations are strongly polarized in liquid or solution spectra. All other bands in liquid or solution are depolarized. Polarization effects are essential to elucidate structures, but are usually ignored in most other applications. Details of the theory and experimental procedure can be found in the literature (15,16).

Raman and infrared bandwidths, expressed in wave numbers, are usually the same. Both spectroscopies have similar strengths and weaknesses for structure elucidation. Both are sensitive to the presence of functional groups and are poor guides to the length of an aliphatic chain. In both cases, environmental effects such as change of solvent cause subtle changes in the spectra, which can be useful. As for the more familiar infrared spectroscopy, Raman spectroscopy has been largely superseded by nmr (see MAGNETIC SPIN RESONANCE) and mass spectrometry (qv) for routine organic structure elucidation.

The intensity of Raman-scattered light increases as the fourth power of the frequency of the incident light. Thus spectra are stronger if excited with uv lasers, but for practical reasons most Raman spectroscopy is performed using green or nir lasers. The majority of incident light is elastically scattered at the same frequency (Rayleigh scatter), but a small number of photons are scattered at a different wavelength. These photons are scattered with a loss of energy (red, or Stokes-shifted Raman) or gain in energy (blue, or anti-Stokes-shifted Raman). The energy shifts correspond to the vibrational energy levels of the molecule as shown in Figure 1. The Raman shift is reported as the wavenumber (cm^{-1}) difference in frequency between the exciting and scattering frequencies, and is therefore independent of the frequency of the light used to generate the spectrum.

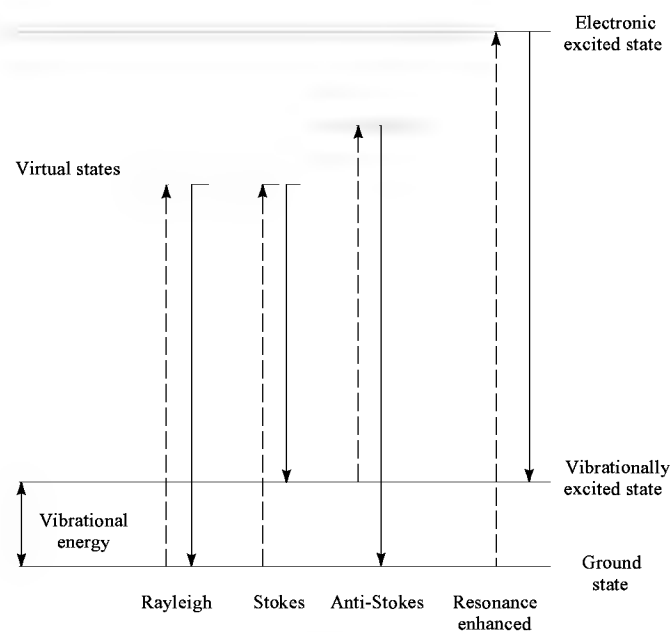


Fig. 1. Schematic of the vibrational energy levels of a molecule where (---) indicate the changes effected by the excitation photon, and (—) those of scattered photons. The energy differences depend on the vibrational levels of the particular scattering molecule.

For most purposes only the Stokes-shifted Raman spectrum, which results from molecules in the ground electronic and vibrational states being excited, is measured and reported. Anti-Stokes spectra arise from molecules in vibrational excited states returning to the ground state. The relative intensities of the Stokes and anti-Stokes bands are proportional to the relative populations of the ground and excited vibrational states. These proportions are temperature-dependent and follow a Boltzmann distribution. At room temperature, the anti-Stokes:Stokes intensity ratio decreases by a factor of 10 with each 480 cm⁻¹ from the exciting frequency. Because of the weakness of the anti-Stokes spectrum (except at low frequency shift), the most important use of this spectrum is for optical temperature measurement (qv) using the Boltzmann distribution function.

Raman scattering is very inefficient compared to either the Rayleigh scattering or fluorescence. For an average scattering liquid, about one in 10³ photons are scattered, but only about one in 10⁴–10⁵ of these scattered photons are Raman-shifted. Because a moderately efficient instrument gathers only 0.1–1% of the scattered photons, the signal is difficult to detect. This drawback is overcome by using a laser to generate the spectrum. A 1-mW green laser emits about 10¹³ photons s⁻¹, resulting in at least 10⁵ Raman-scattered photons s⁻¹. Other tools to increase detection of Raman signal include modern spectrographs which have high throughput, and CCD detectors which have high quantum efficiency and very low dark current. Another significant experimental problem is distinguishing the weak Raman scatter from the much stronger fluorescence and unshifted laser light. This problem, too, can be easily overcome using a well-designed instrument.

Sample Preparation

Ease of Sample Preparation. A significant practical advantage of Raman spectroscopy is the ease of sample handling and preparation. Spectra of solids, liquids, and gases can often be obtained without any sample preparation. A transparent sample is not needed because Raman spectroscopy is a scattering technique. Ordinary quartz or glass containers may be used to hold samples without contributing significantly to background, because both glass and quartz are weak Raman scatterers, and spectra are obtained in the visible–nir regions where quartz and glass are transparent (Fig. 2a). Spectra can also easily be obtained from aqueous solutions, because the water Raman spectrum is weak. Sample volumes on the order of microliters are generally sufficient (Fig. 2b).

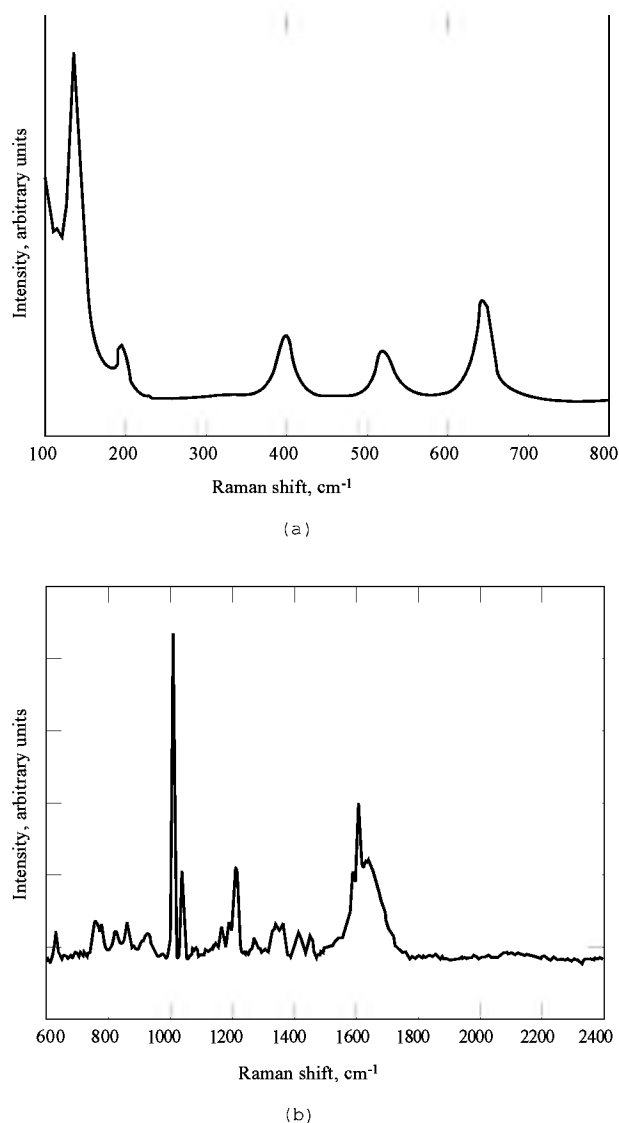


Fig. 2. (a) Raman spectrum of anatase [1317-70-0], TiO₂, in a glass bottle, 532-nm excitation, 0.2-s acquisition. The principal anatase bands are at 143, 195, 392, 512, and 633 cm⁻¹, respectively. (b) Raman spectrum of 10 μL of 0.1-M phenylalanine [63-91-2], 532-nm excitation, 50-s acquisition. The broad peak at 1640 cm⁻¹ is attributed to the O—H bending frequency of H₂O.

Fluorescence Interference. The historical drawback to widespread use of Raman spectroscopy has been the strong fluorescence background exhibited by many materials, even those which are nominally nonfluorescent. This fluorescence often arises from an impurity in the sample, but may be intrinsic to the material being studied. Several methods have proved useful in reducing this background. One of the simplest is sample purification. Another method, called photobleaching, works on robust solids but may cause photodecomposition in many materials. The simplest solution to the fluorescence problem is excitation in the near infrared (750 nm–1.06 μm), where the energy of the incident photons is lower than the electronic transitions of most organic materials, so fluorescence cannot occur. The Raman signal can then be observed more easily. The elimination of fluorescence background more than compensates for the reduction in scattering efficiency in the near infrared. Only in the case of transition-metal compounds, which can fluoresce in the near infrared, is excitation in the midvisible likely to produce superior results in practical samples (17).

Intensity Enhancement Mechanisms

Resonance Raman Spectroscopy. If the excitation wavelength is chosen to correspond to an absorption maximum of the species being studied, a 10²–10⁴ enhancement of the Raman scatter of the chromophore is observed. This effect is called resonance enhancement or resonance Raman (RR) spectroscopy. There are several mechanisms to explain this phenomenon, the most common of which is Franck-Condon enhancement. In this case, a band intensity is enhanced if some component of the vibrational motion is along one of the directions in which the molecule expands in the electronic excited state. The intensity is roughly proportional to the distortion of the molecule along this axis. RR spectroscopy has been an important biochemical tool, and it may have industrial uses in some areas of pigment chemistry. Two biological applications include the determination of helix transitions of deoxyribonucleic acid (DNA) (18), and the elucidation of several peptide structures (19). A review of topics in this area has been published (20).

Surface-Enhanced Raman Spectroscopy. A second technique for increased sensitivity uses the strong enhancement of the electric field of a light wave at certain rough metal surfaces. This surface-enhanced Raman scattering (sers) results in a 10³–10⁴ increase in signal of molecules in contact with the surface. Submonolayers are easily observed, and solution detection limits are 10⁻⁵ to 10⁻⁶ M for ordinary Raman scatterers. The coinage metals, ie, gold, silver, and copper, are most commonly used as the rough surface because their surface electromagnetic states can be excited using visible or near-infrared lasers. The presence of sers has been observed at electrodes, colloids (qv), and metal-island films. Several reviews detailing enhancement theories and recent applications of sers have been published (21).

Instrumentation

In a typical Raman experiment the sample is illuminated using a laser and the scattered light is collected through low f-number optics and focused into a spectrograph or interferometer. Collection at right angles to the exciting beam has been used; however, use of a backscattered geometry is becoming increasingly common.

Modern Raman instruments utilize either a dispersive element (grating) or an interferometer. In most cases the spectrum is analyzed by a grating spectrograph fitted with a CCD array detector. The CCD is a silicon detector, and does not respond to light to the red of 1.05 μm. Therefore, for excitation using a Nd:YAG laser (1.06 μm), a Fabry-Perot interferometer fitted with a germanium or indium gallium arsenide, InGaAs, photodiode is used. Scanning spectrometers with photomultipliers have been obsolete for many years, but are still encountered occasionally (see PHOTODETECTORS).

Lasers having wavelengths ranging from the deep uv to the near infrared have been used in Raman spectroscopy. In industrial laboratories, the most common laser is the Nd:YAG operating at 1.06 μm. Increasingly, diode lasers or other lasers operating in the 750–785-nm region are encountered. These

lasers can be used with spectrograph–CCD instruments which acquire spectra 10–100 times faster than interferometers. In most cases, 750–785 nm is sufficiently red that fluorescence is absent or weak.

The slow-scan CCD, also called the scientific CCD, or in the spectroscopy literature simply CCD, is the detector of choice for most applications of Raman spectroscopy. A well-designed CCD has essentially zero dark current, very low readout noise, and high quantum efficiency (peak 45–70% near 700 nm) in the visible region of the spectrum. However, the response drops quickly above 800 nm and there is no photon response above 1.05 μm. For routine spectroscopy or process control, thermoelectrically cooled (to about –40°C) CCDs are adequate. Although these detectors are somewhat noisier than detectors operated at –100°C or lower, the former do not require liquid nitrogen cooling. The general properties and spectroscopic applications of the CCD have been reviewed (22).

Dispersive Spectrographs. During the 1980s, triple-grating spectrographs were the preferred instruments for Raman spectroscopy where multichannel detection was employed. The first two gratings formed a filter to reject unshifted laser light (Rayleigh scatter), whereas dispersion was provided by the final grating. As of the 1990s, these instruments are employed in some applications because they can be used with any laser wavelength and have excellent rejection of intense Rayleigh scatter at very low Raman shifts. However, the instruments are bulky, expensive, and have poor throughput.

The preferred instrument is a single-stage spectrograph preceded by a Rayleigh line rejection filter. Holographic (Bragg diffraction) filters are most commonly employed, because these offer high (70–80%) throughput and can operate as close as 50 cm^{–1} from the exciting line (see HOLOGRAPHY). For Raman shifts further than a few hundred wavenumbers, the less expensive dielectric filters can be used. The single-stage systems are smaller, simpler, and less expensive than triple spectrographs and have the advantage of higher throughput and better signal-collection efficiency. Most are 0.25–0.5-m Czerny–Turner instruments, but other designs such as the echelle or axial transmissive spectrograph are becoming increasingly popular as of 1994 (23,24). An example of a single-stage Raman spectrograph based on a volume holographic grating and holographic notch filters is shown in Figure 3.

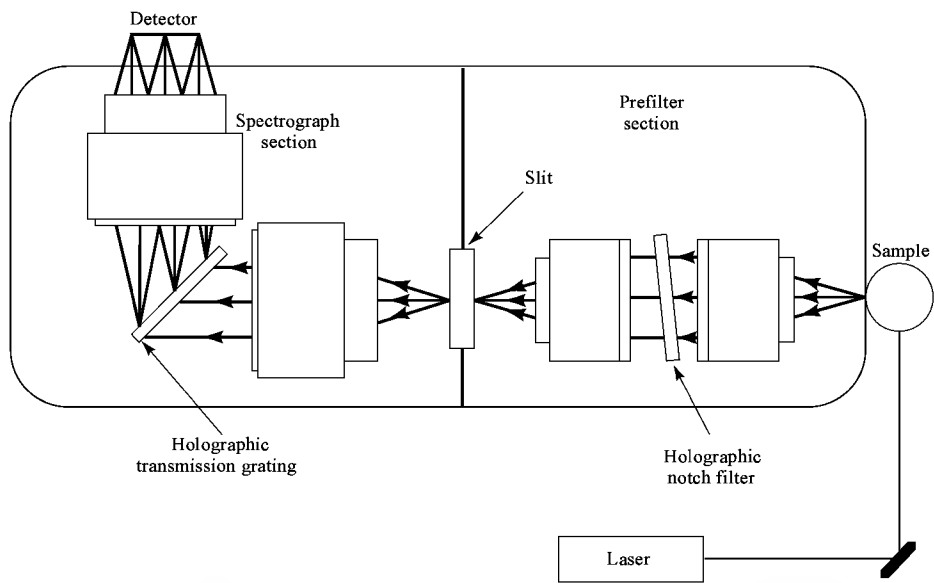


Fig. 3. Schematic diagram of an axial transmissive single-stage Raman spectrometer.

Courtesy of Kaiser Optical Systems, Inc.

The high performance of modern spectrographs means that low power lasers can be used as excitation sources. These are typically 10–100-mW devices which are air-cooled and can be operated from 117-V a-c lines. In the green, the Ar⁺ (514.5-nm) laser remains the most popular but is being challenged by the smaller and more efficient frequency doubled Nd:YAG (532-nm). In the nir, diode lasers (784-nm) and diode-pumped alexandrite [12252-02-7], Al₂BeO₄ (750-nm) lasers are used.

Fourier-Transform Raman Spectroscopy. The first and most popular form of near-infrared Raman spectroscopy employs modified Fourier-transform infrared (ftir) spectrometers, which allow use of Nd:YAG 1064-nm excitation. Using Fourier-transform Raman spectroscopy (ft-Raman) with nir excitation, fluorescence is largely absent in organic materials and Raman spectra are easily obtained. Spectacular early successes included facile acquisition of spectra of nylon and Rhodamine 6G, which has a fluorescence quantum efficiency for green excitation above 90%. An example of a typical ft-Raman spectrometer is shown in Figure 4. The ft-Raman spectrum of the highly fluorescent caffeine derivative theophylline-7-COOH is shown in Figure 5. Although ft-Raman instrumentation has been refined by the vendors, the basic principles remain largely unchanged from the earliest instrument described (25).

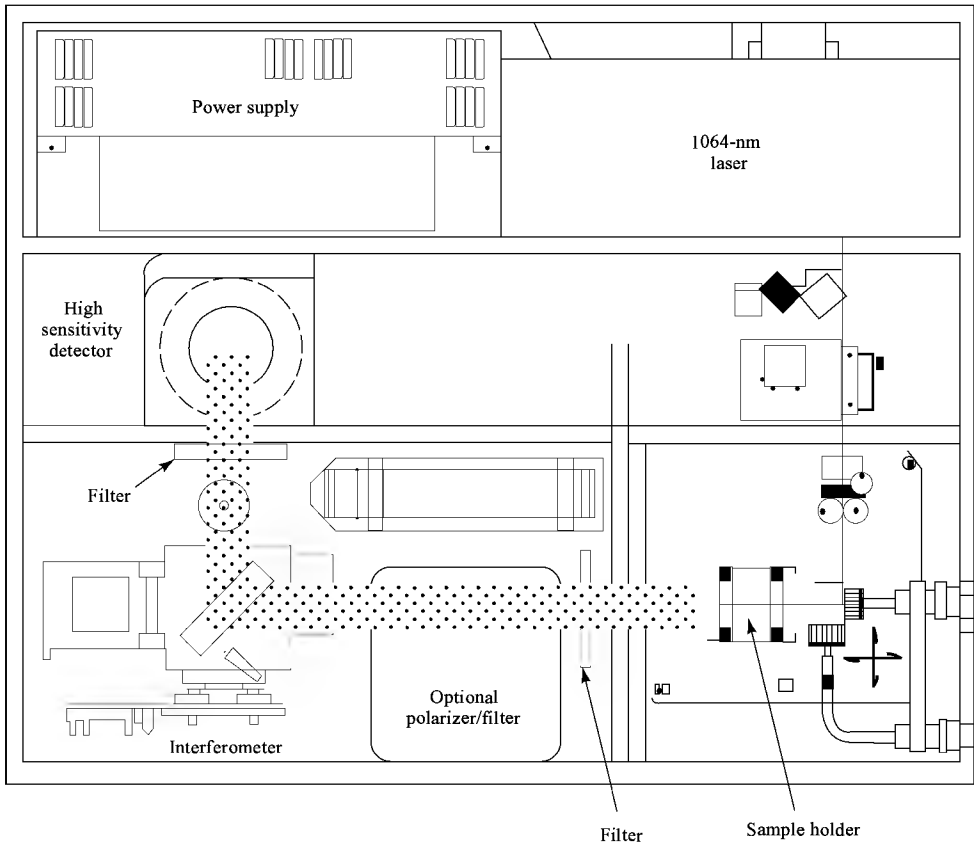


Fig. 4. Schematic diagram of a typical ft-Raman spectrometer.

Courtesy of Nicolet Instrument Corp.

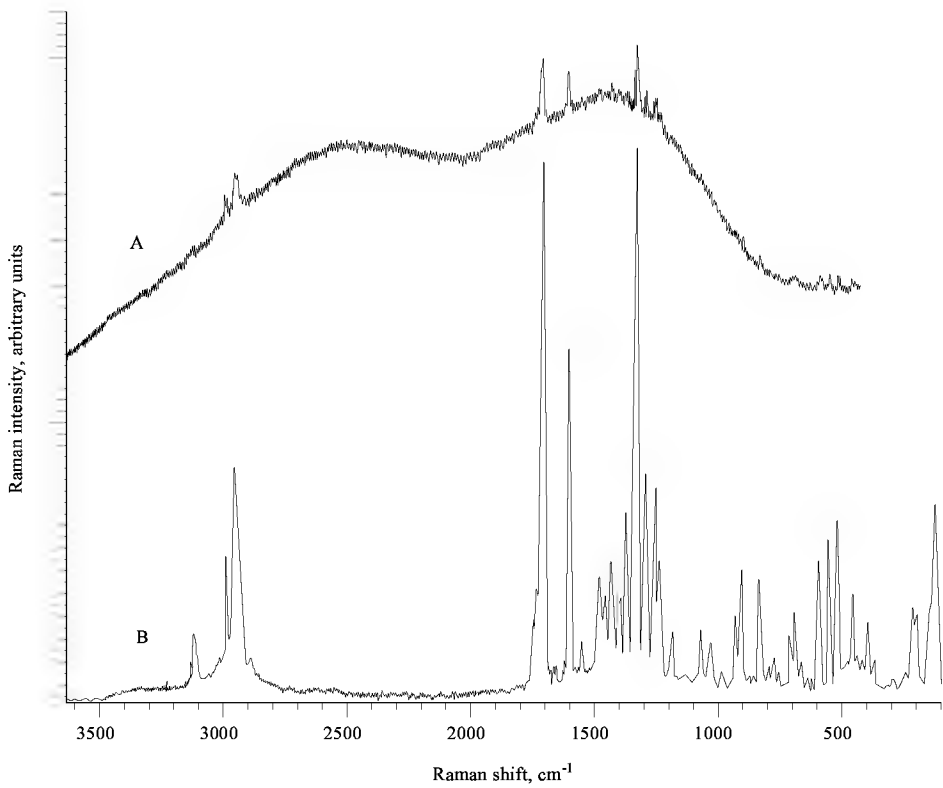


Fig. 5. Raman spectrum of theophylline-7-COOH where A corresponds to standard instrumentation using 514-nm excitation and B to Raman using 1064-nm excitation.

Courtesy of Nicolet Instrument Corp.

Ft-Raman spectra can be acquired for most samples. There have been numerous reports of practical examples, and a journal issue is devoted annually to the topic (26–28). However, the technique does have drawbacks. Most importantly, the low scattering efficiency in the nir and the poor noise performance of nir detectors means that spectrum acquisition is slow (1–10 min). To circumvent this problem, high (>1 W) power lasers are frequently used, but these can cause sample heating, which generates a thermal background that can be just as high as a fluorescence background. A useful review has been published on the advantages and disadvantages of ft-Raman (29) compared to dispersive Raman spectroscopy in the near infrared. Industrial support with a base of installed instruments, as well as the very real fluorescence rejection advantage in some applications, is expected to give ft-Raman a place in the Raman spectroscopy repertoire.

Raman Microspectroscopy. Raman spectra of small solids or small regions of solids can be obtained at a spatial resolution of about 1 μm using a Raman microprobe. A widespread application is in the characterization of materials. For example, the Raman microprobe is used to measure lattice strain in semiconductors (30) and polymers (31,32), and to identify graphitic regions in diamond films (33). The microprobe has long been employed to identify fluid inclusions in minerals (34), and is increasingly popular for identification of inclusions in glass (qv) (35).

A schematic diagram of a typical Raman microprobe is shown in Figure 6. The instrument is a modified epi-illumination fluorescence microscope, or a metallographic microscope, in which the conventional quartz halogen lamp can be replaced by a laser. The laser can be focused to a diameter of $<1\ \mu\text{m}$, depending on the microscope objective used. The backscattered light is gathered through the objective and transferred to the entrance slit of a spectrograph (see MICROSCOPY).

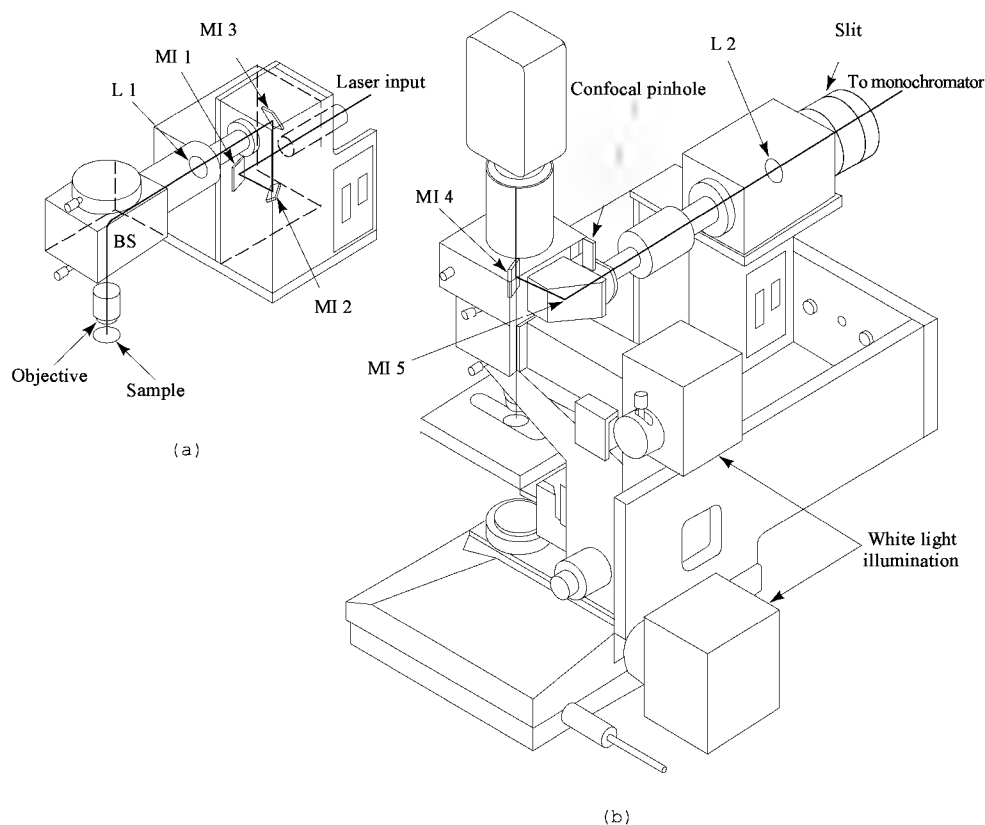


Fig. 6. Schematic diagram of a Raman microprobe where BS = beam splitter, L = lens, and MI = mirror. (a) The illumination pathway; (b) the collection path.

Courtesy of Instruments SA, Inc.

A principal advantage of the Raman microprobe is that the optics are those of a conventional light microscope; a wide variety of special-purpose objectives developed for materials and biological microscopy are available. The Raman microprobe also offers the advantage of fluorescence reduction owing to the high spatial resolution of the microscope if a region of low fluorescence can be chosen for observation.

However, especially using green illumination, sample fluorescence can still be a problem. In some cases the tightly focused laser can burn away fluorescent impurities. Photobleaching can also be attempted; the microprobe can accomplish this in a much shorter time than in a conventional system. NIR excitation can, of course, be used to avoid many fluorescence problems entirely, but the disadvantages of doing so are the limited selection and high prices of microscope objectives designed for operation in the near infrared.

The spatial resolution of the Raman microprobe is about an order of magnitude better than that obtainable using an infrared microscope. Measurement times, typically of a few seconds, are the same as for other Raman spectrographs. To avoid burning samples, low (5–50-mW) power lasers are employed.

Fiber-Optic Probes. Fiber-optic probes provide remote sampling capabilities to Raman instrumentation, are stable, and give reproducible signals. Their historical niche has been in environmental monitoring. More recently these probes have been used in chemical process control and related areas such as incoming materials inspection.

In first-generation designs, a single fiber is used to deliver laser light to the sample, and a bundle of fibers collects the scattered light. These probes suffer from a background which is composed of silica Raman scatter and cladding fluorescence, but are inexpensive to manufacture. The background limits the usable length, and can be especially troublesome if the sample is a highly scattering solid.

More recent probe designs employ dielectric or holographic filters at the distal (sample) end to reduce or largely eliminate probe background. The filters deliver the excitation wavelength to the sample and prevent backscattered laser light from generating fluorescence or Raman scatter in the collection fiber(s). The filters and their mountings increase both the cost and bulk of the probes. The low background of filtered probes makes them preferred in all but the least demanding applications (36).

Uses

Diamond and Graphite. Diamond [7782-40-3] is among the strongest Raman scatterers, and Raman spectroscopy is an important diagnostic tool for determining diamond film quality (37) (see CARBON). Diamond has an intense Raman band at $1332\ \text{cm}^{-1}$, the presence of which verifies that a material contains diamond. The more regular the crystal lattice, the narrower the width of this band. Graphite [7782-42-5] has two first-order phonon Raman bands, a strong band at $1580\ \text{cm}^{-1}$ and a weaker one at $1357\ \text{cm}^{-1}$. There is also a second-order phonon band at $2700\ \text{cm}^{-1}$. The exact positions and widths of these bands are indicative of the type of graphite. Amorphous carbon generates broad ($20\text{--}100\ \text{cm}^{-1}$) bands in the $1100\text{--}1800\ \text{cm}^{-1}$ region. Consequently, the Raman spectrum can be used to identify the presence of graphitic regions in diamond films, or amorphous carbon regions in graphite sheets and fibers (see CARBON AND GRAPHITE FIBERS).

Raman spectroscopy of graphite can be an experimental challenge, because the material is a strong blackbody absorber. Generally, low (1–10-mW) laser power is used to minimize heating, which causes the band positions to change. In addition, the expansion of the graphite causes the material to go out of the focus of the optical system, an effect which can be more pronounced in microprobe work.

Titanium Dioxide and Other Metal Oxides. The crystal forms of titanium dioxide [13463-67-7], TiO_2 , have easily distinguishable Raman spectra. The important pigment, rutile [1317-80-2], has an intense Raman band at $610\ \text{cm}^{-1}$, which is a useful diagnostic. This band is finding increasing application in incoming materials inspection (38) (see PIGMENTS, INORGANIC; TITANIUM COMPOUNDS, INORGANIC). Tungsten and molybdenum catalysts have also been characterized by Raman spectroscopy (39).

Other Inorganics. Inorganic species in solution have been studied very effectively by Raman spectroscopy. Work in this area includes the investigation of coordination compounds (qv) of fluorine (qv) (40), the characterization of low dimensional materials (41) and coordinated ligands (42), and single-crystal studies (43). Several compilations of characteristic vibrational frequencies of main-group elements have been published to aid in the identification of these species (44,45).

Polymers. Infrared and Raman spectroscopy have rather similar strengths and weaknesses for use in polymer characterization. Using near-ir (nir) Raman spectroscopy, the fluorescence of industrial polymers whether through additives or impurities is no longer an impediment. The classic case is nylon [32131-17-2]. The near infrared-excited ft-Raman spectrum is easily acquired, whereas the green-excited spectrum is buried beneath an overwhelming fluorescence. Excitation in the near-ir is not always necessary. Raman spectra of many polymers can be obtained using visible excitation as shown in Figure 7.

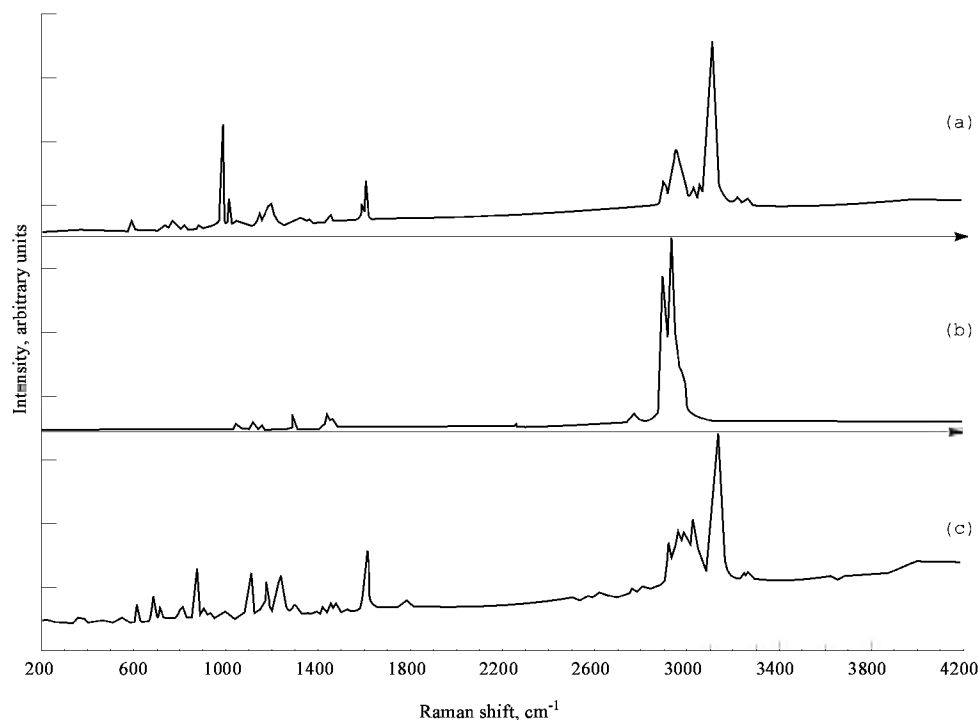


Fig. 7. Raman microprobe spectra of (a) polystyrene [9003-53-6], (b) low density polyethylene, and (c) polycarbonate [24936-68-3].

Courtesy of Instruments SA, Inc.

The ease of sample handling makes Raman spectroscopy increasingly preferred. Like infrared spectroscopy, Raman scattering can be used to identify functional groups commonly found in polymers, including aromaticity, double bonds, and C bond H stretches. More commonly, the Raman spectrum is used to characterize the degree of crystallinity or the orientation of the polymer chains in such structures as tubes, fibers (qv), sheets, powders, and films (see THIN FILMS).

More recently, Raman spectroscopy has been used to investigate the vibrational spectroscopy of polymer liquid crystals (46) (see LIQUID CRYSTALLINE MATERIALS), the kinetics of polymerization (47) (see KINETIC MEASUREMENTS), synthetic polymers and rubbers (48), and stress and strain in fibers and composites (49) (see COMPOSITE MATERIALS). The relationship between Raman spectra and the structure of conjugated and conducting polymers has been reviewed (50,51). In addition, a general review of ft-Raman studies of polymers has been published (52).

Biological Systems. Whereas Raman spectroscopy is an important tool of physical biochemistry, much of this elegant work is of scant interest to the industrial chemist. However, Raman spectroscopy has been used to locate cancerous cells in breast tissue (53) and find cataractous tissue in eye lenses (54), suggesting a role in industrial hygiene (qv). Similarly, the Raman spectra of bacteria are surprisingly characteristic (55) and practical applications are beginning to emerge.

Physical Properties. Raman spectroscopy is an excellent tool for investigating stress and strain in many different materials (see MATERIALS RELIABILITY). Lattice strain distribution measurements in silicon are a classic case. More recent examples of this include the characterization of thin films (56), and measurements of stress and relaxation in silicon-germanium layers (57).

Process Control and Environmental Monitoring. A nascent and potentially important application area for Raman spectroscopy is industrial process control. Concentrations of materials are typically high, and the bands of starting materials and products are usually characteristic. Less reliance on multivariate calibration is needed than is typical for nir-based process control measurements. The development of compact Raman monitoring instrumentation is indicative of serious industrial interest. Some studies have been reported, including demonstrations to distillation (qv) process control (58) and styrene polymerization (59). Similarly, instrumentation developments have made environmental monitoring (60), oxygen (qv) detection during combustion (61), and remote monitoring and detection of toxic materials practical.

Raman spectroscopy is a rapidly expanding field where new applications appear regularly, often in areas where the technique was previously considered inapplicable. Driven by the needs of chemical process control, environmental monitoring, and biomedical diagnostics, instrument vendors are developing specialized instruments and probes optimally suited for these fields. Using this instrumentation, industrial chemists can now exploit the familiar Raman spectroscopic advantages of high resolution, easy remote sampling, and ease of application to both solids and aqueous solutions.

BIBLIOGRAPHY

1. C. V. Raman and K. S. Krishnan, *Nature* **121**, 501 (1928).
2. J. G. Grasselli, F. Walder, C. Petty, and G. Kamany, *J. Mol. Spectrosc.* **294**, 207-210 (1993).
3. N. J. Everall and J. Lumsdon, *Spectrosc. Eur.* **4**, 10-21 (1992).
4. B. Prevot and J. Wagner, *J. Prog. Cryst. Growth Charact. Mater.* **22**, 245-319 (1992).
5. E. Liarokapis, *NATO ASI Ser., Ser. B* **286**, 457-472 (1991).
6. C. G. Smith and co-workers, *Anal. Chem.* **65**, 217-243 (1993).
7. G. Ellis, C. Marco, and F. J. M. Gomez, *Rev. Plast. Mod.* **63**, 181-187 (1992).
8. S. Farquharson and S. F. Simpson, *Proc. SPIE Int. Soc. Opt. Eng.* **1681**, 276-290 (1992).
9. S. E. Nave, P. E. O'Rourke, R. A. Malstrom, and W. S. Prather, *Process Control Qual.* **3**, 43-48 (1992).
10. R. J. H. Clark and R. E. Hester, eds., *Spectroscopy of Biological Systems*, John Wiley & Sons, Inc., New York, 1986.
11. A. J. P. Alix, L. Bernard, and M. Manfait, eds., *Spectroscopy of Biological Molecules*, John Wiley & Sons, Ltd., Chichester, U.K., 1985.
12. P. R. Carey, ed., *Biochemical Applications of Raman and Resonance Raman Spectroscopies*, Academic Press, Inc., New York, 1982.
13. D. L. Gerrard and J. Birnie, *Anal. Chem.* **64**, 502R-513R (1992).
14. D. L. Gerrard, *Anal. Chem.* **66**, 547R-557R (1994).
15. D. A. Long, *Raman Spectroscopy*, McGraw Hill, New York, 1977.

16. J. G. Grasselli and B. J. Bulkin, eds., *Analytical Raman Spectroscopy*, John Wiley & Sons, Inc., New York, 1991.

17. B. Schrader, A. Hoffmann, and S. Keller, *Spectrochim. Acta, Part A* **47A**, 1135–1148 (1991).

18. P. Miskovsky, A. Laigle, L. Chinsky, and P. Y. Turpin, *J. Biomol. Struct. Dyn.* **10**, 169–179 (1992).

19. Y. Wang, R. Purrello, T. Jordan, and T. G. Spiro, *J. Amer. Chem. Soc.* **113**, 6359–6368 (1991).

20. S. A. Asher, *Anal. Chem.*, **65**, 59A–66A and 201A–210A (1993).

21. R. K. Chang and T. E. Furtak, ed., *Surface Enhanced Raman Scattering*, Plenum Press, New York, 1982.

22. P. M. Epperson, J. V. Sweedler, R. B. Bilhorn, G. R. Sims, and M. B. Denton, *Anal. Chem.* **60**, 327A–335A (1988).

23. D. E. Battey, J. B. Slater, R. Wludyka, H. Owen, D. M. Pallister, and M. D. Morris, *Appl. Spectrosc.* **47**, 1913–1919 (1993).

24. M. J. Pelletier, *Appl. Spectrosc.* **44**, 1699–1705 (1990).

25. T. Hirschfeld and B. Chase, *Appl. Spectrosc.* **40**, 133 (1986).

26. *Spectrochim. Acta, Part A* **49A**(5–6) (1993).

27. *Spectrochim. Acta, Part A* **47A**(9–10) (1991).

28. *Spectrochim. Acta, Part A* **46A**(2) (1990).

29. S. F. Parker, N. Conroy, and V. Patel, *Spectrochim. Acta, Part A* **49A**, 657–666 (1993).

30. S. Nakashima, *NATO ASI Ser., Ser. B* **273**, 291–309 (1991).

31. R. Tabaksblat, R. J. Meier, and B. J. Kip, *J. Appl. Spectrosc.* **46**, 60–68 (1992).

32. J. M. Paster, *Makromol. Chem. Macromol. Symp.* **52**, 57–73 (1991).

33. P. R. Chalker, *NATO ASI Ser., Ser. B* **266**, 127–150 (1991).

34. R. W. T. Wilkins and L. Jenatton, *J. Geochem. Explor.* **42**, 157–166 (1991).

35. J. S. Sanghera and I. D. Aggarwal, *J. Amer. Ceram. Soc.* **76**, 2341–2344 (1993).

36. S. K. Sharma, C. L. Schoen, and T. F. Cooney, *Appl. Spectrosc.* **47**, 377–379 (1993).

37. D. S. Knight and W. B. White, *J. Mater. Res.* **4**, 385–393 (1989).

38. R. J. Capwell, F. Spagnolo, and M. A. DeSesa, *Appl. Spectrosc.* **26**, 537–539 (1972).

39. M. Mehicic and J. G. Grasselli, in Ref. 16, p. 325.

40. D. J. Reynolds, *Advan. Fluor. Chem.* **7**, 1–68 (1973).

41. S. Nakashima and M. Hangyo, *Croat. Chem. Acta* **61**, 331–346 (1988).

42. G. Davidson, *Prop. Inorg. Organomet. Comp.* **14**, 236–302 (1981).

43. D. M. Adams, and P. D. Halton, *Spectrosc. Prop. Inorg. Organomet. Comp.* **14**, 173–197 (1981).

44. G. Davidson, *Spectrosc. Prop. Inorg. Organomet. Comp.* **18**, 220–250 (1985).

45. S. Cradock, *Spectrosc. Prop. Inorg. Organomet. Comp.* **14**, 198–212 (1981).

46. M. P. Fontana and co-workers, *Mol. Cryst. Liq. Cryst. Sci. Technol., Sect. A* **243**, 31–49 (1994).

47. H. G. M. Edwards, A. F. Johnson, and I. R. Lewis, *J. Raman Spectrosc.* **24**, 475–483 (1993).

48. C. G. Smith and co-workers, *Anal. Chem.* **65**, 217r–243r (1993).

49. C. Filiou, C. Galiotis, and D. N. Batchelder, *Composites* **23**, 28–38 (1992).

50. Y. Yacoby and E. Ehrenfreund, *Top. Appl. Phys.* **68**, 73–135 (1991).

51. H. G. M. Edwards, A. F. Johnson, and I. R. Lewis, *J. Raman Spectrosc.* **24**, 475–483 (1993).

52. P. J. Hendra and C. H. Jones, *Makromol. Chem. Macromol. Symp.* **52**, 41–56 (1991).

53. D. C. Redd, Z. C. Feng, K. T. Yue, and T. S. Gansler, *Appl. Spectrosc.* **47**, 787–791 (1993).

54. K. V. Sokolov, S. V. Lutsenko, I. R. Nabiev, S. Nie, and N. T. Yu, *Appl. Spectrosc.* **45**, 1143–1148 (1991).

55. S. Chadha, W. H. Nelson, and J. F. Sperry, *Rev. Sci. Instrum.* **64**, 3088–3093 (1993).

56. V. A. Yakolev, *Laser Focus World* **29**, 83–89 (1993).

57. B. Dietrich and co-workers, *J. Appl. Phys.* **74**, 3177–3180 (1993).

58. M. J. Roberts and co-workers, *Adv. Instrum. Control* **45**(P1), 463–468 (1990).

59. S. M. Angel, T. J. Kulp, T. M. Vess, *Appl. Spectrosc.* **46**, 1085–1091 (1992).

60. M. M. Carrabba, K. M. Spencer, and R. D. Rauk, *Proc. SPIE-Int. Soc. Opt. Eng.* **1434**, 1127–1134 (1991).

61. A. Leipertz, *Appl. Opt.* **22**, 901–903 (1983).

Kenneth A. Christensen
Elizabeth A. Todd
Michael D. Morris
University of Michigan

INITIATORS

Free-radical initiators,
Anionic initiators,
Cationic initiators,

FREE-RADICAL INITIATORS

Free-radical initiators are chemical substances that, under certain conditions, initiate chemical reactions by producing free radicals:



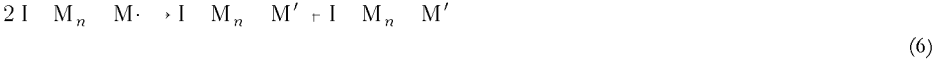
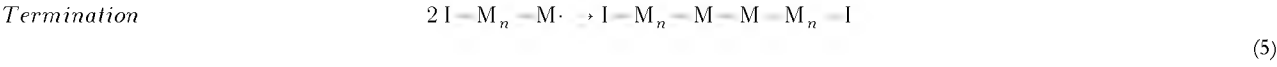
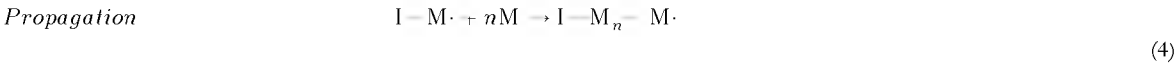
Initiators contain one or more labile bonds that cleave homolytically when sufficient energy is supplied to the molecule. The energy must be greater than the bond dissociation energy (BDE) of the labile bond. Radicals are reactive chemical species possessing a free (unbonded or unpaired) electron. Radicals may also be positively or negatively charged species carrying a free electron (ion radicals). Initiator-derived radicals are very reactive chemical intermediates and generally have short lifetimes, ie, half-life times less than 10⁻³ seconds (1).

The principal commercial initiators used to generate radicals are peroxides and azo compounds. Lesser amounts of carbon–carbon initiators and photoinitiators, and high energy ionizing radiation are also employed commercially to generate radicals.

There are three general processes for supplying the energy necessary to generate radicals from initiators: thermal processes, microwave or ultraviolet (uv) radiation processes, and electron transfer (redox) processes. Radicals can also be produced in high energy radiation processes. Initiators are sometimes called radical catalysts. However, initiators are not true catalysts because they are consumed in amounts ranging from substoichiometric up to stoichiometric or greater when they are employed as initiators in chemical reactions. True catalysts such as enzymes are not consumed in the chemical reaction that they catalyze.

Once formed, radicals undergo two basic types of reactions: propagation reactions and termination reactions. In a propagation reaction, a radical reacts to form a covalent bond and to generate a new radical. The three most common propagating reactions are atom abstraction, β-scission, and addition to carbon–carbon double bonds or aromatic rings. In a termination reaction, two radicals interact in a mutually destructive reaction in which both radicals form covalent bonds and reaction ceases. The two most common termination reactions are coupling and disproportionation. Because the propagation reaction is a chain reaction, it has become the most significant aspect of commercial free-radical chemistry. Radical chain reactions are involved in many commercial processes.

Radicals are employed widely in the polymer industry, where their chain-propagating behavior transforms vinyl monomers into polymers and copolymers. The mechanism of addition polymerization involves all three types of reactions discussed above, ie, initiation, propagation by addition to carbon–carbon double bonds, and termination:



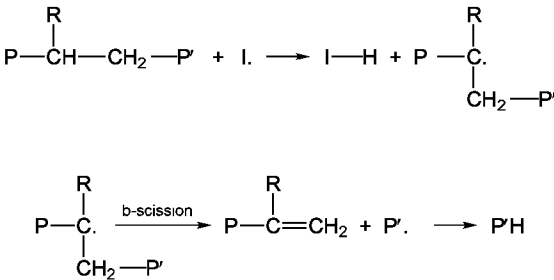
In these equations I is the initiator and I· is the radical intermediate, M is a vinyl monomer, I–M· is an initial monomer radical, I–M_nM· is a propagating polymer radical, and M' and M' are polymer end groups that result from termination by disproportionation. Common vinyl monomers that can be homo- or copolymerized by radical initiation include ethylene, butadiene, styrene, vinyl chloride, vinyl acetate, acrylic and methacrylic acid esters, acrylonitrile, N-vinylimidazole, N-vinyl-2-pyrrolidinone, and others (2).

Two other important commercial uses of initiators are in polymer cross-linking and polymer degradation. In a cross-linking reaction, atom abstraction, usually a hydrogen abstraction, occurs, followed by termination by coupling of two polymer radicals to form a covalent cross-link:



P–H is a polymer with covalently attached hydrogen, I· is the initiating radical, and P–P is a cross-linked polymer. Cross-linking is a commercially

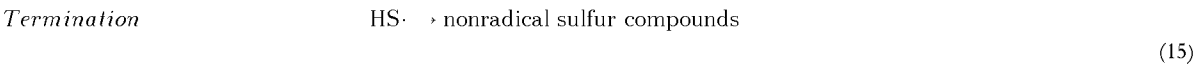
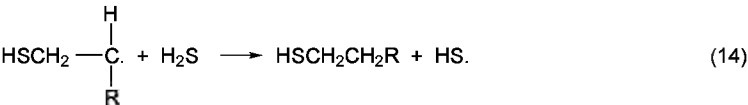
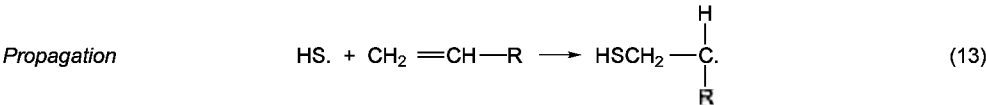
important reaction of thermoplastics (such as polyethylene) and elastomers. In polymer degradation, hydrogen abstraction is followed by β -scission that results in breakage of the polymer chain:



$\text{I}\cdot$ is the initiating radical, $\text{P}'\cdot$ is the chain-propagating polymer radical that subsequently abstracts a hydrogen atom from another polymer molecule, $\text{P}-\text{CHR}-\text{CH}_2-\text{P}'$ is the polymer before, and $\text{P}-\text{CR}=\text{CH}_2$ and $\text{P}'\text{H}$ are polymer chains after degradation. Polymer degradation is important in facilitating the commercial processing (molding and extruding) of polypropylene (the degradation is more commonly called controlled rheology or vis-breaking). In the β -scission reaction the first-formed radical cleaves to a polymer radical and to an electronically neutral molecule (polymer with an unsaturated end group) by scission of a carbon–carbon bond β to the atom bearing the initial radical center.

Other common radical-initiated polymer processes include curing of resins, eg, unsaturated polyester–styrene blends; curing of rubber; grafting of vinyl monomers onto polymer backbones; and telomerizations.

A typical example of a nonpolymeric chain-propagating radical reaction is the anti-Markovnikov addition of hydrogen sulfide to a terminal olefin. The mechanism involves alternating abstraction and addition reactions in the propagating steps:



Other nonpolymeric radical-initiated processes include oxidation, autoxidation of hydrocarbons, chlorination, bromination, and other additions to double bonds. The same types of initiators are generally used for initiating polymerization and nonpolymerization reactions. Radical reactions are extensively discussed in the chemical literature (3–15).

Structure–Reactivity Relationships. Much has been written about the structure–reactivity of radicals. No single unifying concept has satisfactorily explained all radical reactions reported in the literature. A longstanding correlation of structure and reactivity involves comparisons of the energies required to homolytically break covalent bonds to hydrogen. It is assumed that this energy, the hydrogen bond dissociation energy (BDE), reflects the stability and the reactivity of the radical coproduced with the hydrogen atom (16). However, this assumption should really be limited to radical reactivity and selectivity in hydrogen atom abstraction reactions, and can be particularly misleading for reactions with polar transition states, in which radicals can behave either as nucleophiles or electrophiles (17). Nevertheless, the correlation of radical reactivity with BDE is quite useful. Table 1 shows some general BDE values for the formation of various carbon and oxygen radicals from various precursors. According to the theory, the higher the BDE, the higher the reactivity and the lower the stability of the radical formed by removal of a hydrogen atom. Thus bulky *tert*-alkyl radicals are more stable and less reactive than less bulky secondary alkyl radicals that in turn are more stable and less reactive than primary alkyl radicals. Hydroxyl radicals are the most reactive radicals listed. Methyl radicals are more reactive than other primary alkyl radicals and are about as reactive as alkoxy radicals. Lower stability and increased reactivity correspond to less discriminating radical behavior, resulting in faster and less selective radical reactions with other molecules. In organic systems, this reaction is usually hydrogen atom abstraction. Consequently, methyl radicals and oxy radicals are considered good hydrogen atom-abstracting radicals, and are suitable for cross-linking, grafting, and degradation reactions. Enhanced stability and reduced reactivity correspond to more discriminating radical behavior, resulting in slower and more selective subsequent reactions. Therefore, reactions other than hydrogen abstraction are favored. Substituted carbon radicals, such as the ethyl radical, are ineffective hydrogen-abstracting radicals; thus these radicals are more likely to add to carbon–carbon double bonds. Initiators that generate these types of radicals are suitable for vinyl monomer polymerizations that avoid undesirable side reactions (cross-linking, grafting, etc).

Table 1. Bond Dissociation Energies

Precursor	BDE, kJ mol ^a
(R) ₃ C–H	381
(R) ₂ CH–H	406
RCH ₂ –H	418
CH ₃ –H	439
RO–H	439
RCOO–H	444
C ₆ H ₅ –H	469
HO–H	498

^a To convert kJ mol to kcal mol, divide by 4.184.

The BDE theory does not explain all observed experimental results. Addition reactions are not adequately handled at all, mostly owing to steric and electronic effects in the transition state. Thus it is important to consider both the reactivities of the radical and the intended coreactant or environment in any attempt to predict the course of a radical reaction (18). Application of frontier molecular orbital theory may be more appropriate to explain certain reactions (19).

The choice of an initiator for a given radical process depends on the reaction conditions and the reactivity of the initiator. These two factors must be balanced so that the reaction is successful. Knowing the decomposition behavior of initiators is important to proper selection. The stabilities or reactivities of initiators such as organic peroxides and aliphatic azo compounds are significantly affected by structural variations close to the labile bond or bonds, ie, the oxygen–oxygen bond in peroxides and the carbon–nitrogen bonds in aliphatic azo compounds. The reactivity differences, resulting from structural differences between initiators, are due to several electronic and steric factors. Alkyl and aryl substituents stabilize carbon radicals through resonance and field effects. These substituent effects on radical stability are reversed for initiator stability. Initiators that decompose to produce highly alkylated or arylated carbon radicals are less stable (more reactive) than those that decompose to less alkylated or arylated carbon radicals. Electronic factors introduced by electron-donating or electron-withdrawing substituents can also affect initiator stability–reactivity; electron-donating substituents stabilize, whereas electron-withdrawing groups destabilize incipient carbon radicals. Initiators with bulky groups on either side of the labile, radical-forming bonds are less stable (more reactive) than initiators with less bulky groups since decomposition to radicals relieves ground-state steric strain.

Activation Parameters. Thermal processes are commonly used to break labile initiator bonds in order to form radicals. The amount of thermal energy necessary varies with the environment, but absolute temperature, *T*, is usually the dominant factor. The energy barrier, the minimum amount of energy that must be supplied, is called the activation energy, *E_a*. A third important factor, known as the frequency factor, *A*, is a measure of bond motion freedom (translational, rotational, and vibrational) in the activated complex or transition state. The relationships of *A*, *E_a*, and *T* to the initiator decomposition rate (*k_d*) are expressed by the Arrhenius first-order rate equation (eq. 16) where *R* is the gas constant, and *A* and *E_a* are known as the activation parameters.

$$k_d = Ae^{(-E_a/RT)} \text{ or } \ln k_d = \ln(A) - E_a/RT \tag{16}$$

Increasing temperature increases initiator decomposition rate. When a single labile bond is broken in the rate-determining step, the frequency factor is high. When multiple bonds are broken, the activated complex is restricted, the frequency factor is low, and the rate of decomposition is reduced (assuming no change in activation energy). Generally, slower rates of decomposition of the initiator mean higher activation energy values. Steric and electronic factors affect the activation energy of the initiator. Factors that enhance the stabilities of the incipient radicals reduce the activation energy and thus increase the decomposition rate.

The activation parameters for an initiator can be determined at normal atmospheric pressure by plotting $\ln k_d$ vs $1/T$ using initiator decomposition rates obtained in dilute solution (0.2 *M* or lower) at several temperatures. Rate data from dilute solutions are required in order to avoid higher order reactions such as induced decompositions. The intercept for the resulting straight line is $\ln A$ and the slope of the line is $-E_a/R$, therefore both *A* and *E_a* can be calculated.

Half-Life. Once these activation parameters have been determined for an initiator, half-life times at a given temperature, ie, the time required for 50% decomposition at a selected temperature, and half-life temperatures for a given period, ie, the temperature required for 50% decomposition of an initiator over a given time, can be calculated. In selecting appropriate initiators for radical applications such as vinyl monomer polymerizations and polyolefin cross-linking, care must be exercised in the use of calculated half-life data for temperatures, pressures, and solvents different than those used in determining the activation parameters. Half-life data are useful for comparing the activity of one initiator with another when the half-life data are determined in the same solvent and at the same concentration and, preferably, when the initiators are of the same class. Because producers of initiators and their customers roughly correlate the thermal stability of initiators with temperature, it is useful to express this stability in terms of 1 and 10-h half-life temperatures, ie, the temperatures at which 50% of the initiator has decomposed in 1 and 10 h, respectively. An extensive compilation of rate data for initiators is available (20). Rate data for commercial organic peroxide initiators are summarized in a half-life bulletin (21).

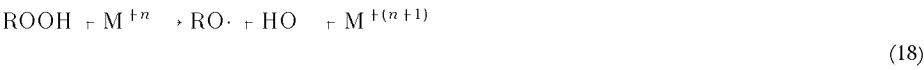
Although a variety of methods for generating radicals by one or more of these three methods are reported in the literature, commercial initiators are primarily organic and inorganic peroxides, aliphatic azo compounds, certain organic compounds with labile carbon–carbon bonds, and photoinitiators.

Organic Peroxides

Organic peroxides are compounds possessing one or more oxygen–oxygen bonds. They have the general structure ROOR' or ROOH, and decompose thermally by the initial cleavage of the oxygen–oxygen bond to produce two radicals:



Depending on the peroxide class, the rates of decomposition of organic peroxides can be enhanced by specific promoters or activators, which significantly decrease the energy necessary to break the oxygen–oxygen bond. Such accelerated decompositions occur well below the peroxides' normal application temperatures and usually result in generation of only one useful radical, instead of two. An example is the decomposition of hydroperoxides with multivalent metals (M), commonly iron, cobalt, or vanadium:



Solvent polarity also affects the rate of peroxide decomposition. Most peroxides decompose faster in more polar or polarizable solvents. This is true even if the peroxide is not generally susceptible to higher order decomposition reactions. This phenomenon is illustrated by various half-life data for *tert*-butyl peroxyphthalate [927-07-1]. The 10-h half-life temperature for *tert*-butyl peroxyphthalate varies from 62°C in decane (nonpolar) to 55°C in benzene (polarizable) and 53°C in methanol (polar).

Following radical generation, the radicals produced (RO· and R'O·) can initiate the desired reaction. However, when the radicals are generated in commercial applications, they are surrounded by a solvent, monomer, or polymer "cage." When the cage is solvent, the radical must diffuse out of this cage to react with the desired substrate. When the cage is monomer, the radical can react with the cage wall or diffuse out of the cage. When the cage is polymer, reaction with the polymer can occur in the cage. Unfortunately, other reactions can occur within the cage and can adversely affect efficiency of radical generation and radical reactivity. If the solvent reacts with the initiator radical, then solvent radicals may participate in the desired reaction.

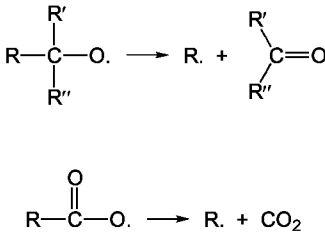
Two secondary propagating reactions often accompany the initial peroxide decomposition: radical-induced decompositions and β-scission reactions. Both reactions affect the reactivity and efficiency of the initiation process. Peroxydicarbonates and hydroperoxides are particularly susceptible to radical-induced decompositions. In radical-induced decomposition, a radical in the system reacts with undecomposed peroxide, eg:



Radical-induced decomposition is an inefficient method of generation of radicals, since the peroxide decomposes without adding radicals to the system. Such decompositions are suppressed in vinyl monomer polymerizations, since the vinyl monomers quickly and efficiently scavenge radicals. In nonscavenging environments, eg, in nonolefinic solvents, induced decompositions occur with those peroxides that are susceptible, and they become more pronounced as the peroxide concentration increases. Whereas the homolysis of organic peroxides is a first-order reaction, the radical-induced decomposition is generally a higher order reaction. Therefore, in those peroxide systems where induced decomposition is occurring, decomposition rates

are significantly higher than the true first-order decomposition rates.

The other secondary propagation reaction that occurs during initiation is β -scission as shown in equations 20 and 21:



Although reaction 21 is a β -scission reaction, it is more commonly termed decarboxylation. In both reactions, the energetics and other properties of the radicals are changed. The initially formed oxygen radicals become carbon radicals. The earlier discussion of relative BDEs for the two types of radicals is applicable here. Steric and temperature effects are also important in β -scission. In equation 20, the newly formed alkyl radical R $\cdot$ is generally derived from the bulkiest alkyl group of the alkoxy radical since it is usually the most stable radical. An exception here is the phenyl radical that does not form upon β -scission of α -cumyloxy radicals, owing to its high energy. Instead, β -scission of the α -cumyloxy radical gives methyl radical and acetophenone. For *tert*-alkoxy radicals, the difference in generation of a *tert*-butoxy radical or a *tert*-amyloxy radical from a *tert*-alkyl peroxide can make a significant difference in the course of the resulting radical reaction. β -Scission of the *tert*-butoxy radical produces a methyl radical having about the same energy (as indicated by BDE) and reactivity as a *tert*-butoxy radical. β -Scission of the *tert*-amyloxy radical produces an ethyl radical having significantly lower energy and reactivity than the *tert*-amyloxy radical.

In equation 21, only one alkyl radical is possible; however, the rate of β -scission is greatly influenced by the bulk of the R group. As with β -scission of α -cumyloxy radicals, benzoyloxy radicals do not decarboxylate as readily as other acyloxy radicals, owing to formation of high energy phenyl radicals. If the R group is sufficiently bulky, decarboxylation occurs simultaneously with scission of the oxygen–oxygen bond. Increased temperatures enhance β -scission. For more thermally stable peroxides, the higher decomposition temperatures result in increased β -scission.

Approximately 100 different organic peroxide initiators, in well over 300 formulations, are commercially produced throughout the world, primarily for the polymer and resin industries (22–27). A multidient study covers the commercial producers and users of organic peroxides as well as other initiators, and their commercial markets and applications (28).

The eight classes of organic peroxides that are produced commercially for use as initiators are listed in Table 2. Included are the 10-h half-life temperature ranges (nonpromoted) for the members of each peroxide class.

Table 2. Commercial Organic Peroxide Classes

Organic peroxide class	Structure ^a	10-h $t_{1/2}$ ^{b,c} , °C
diacyl peroxides	$\begin{array}{c} \text{O} \qquad \qquad \text{O} \\ \qquad \qquad \\ \text{R}-\text{C}-\text{OO}-\text{C}-\text{R} \end{array}$	21–75
dialkyl peroxydicarbonates	$\begin{array}{c} \text{O} \qquad \qquad \text{O} \\ \qquad \qquad \\ \text{RO}-\text{C}-\text{OO}-\text{C}-\text{OR} \end{array}$	49–51
<i>tert</i> -alkyl peroxyesters	$\begin{array}{c} \text{O} \\ \\ \text{R}-\text{C}-\text{OO}-t\text{-R} \end{array}$	38–107
<i>OO-tert</i> -alkyl <i>O</i> -alkyl monoperoxycarbonates	$\begin{array}{c} \text{O} \\ \\ \text{RO}-\text{C}-\text{OO}-t\text{-R} \end{array}$	99–100
di(<i>tert</i> -alkylperoxy)ketals	$\begin{array}{c} \text{R}' \qquad \text{OO}-t\text{-R} \\ \diagdown \quad / \\ \text{C} \\ / \quad \diagdown \\ \text{R} \qquad \text{OO}-t\text{-R} \end{array}$	92–110
di- <i>tert</i> -alkyl peroxides	$t\text{-R}-\text{OO}-t\text{-R}$	115–128
<i>tert</i> -alkyl hydroperoxides	$t\text{-R}-\text{OO}-\text{H}$	not applicable
ketone peroxides	$\begin{array}{c} \text{R}' \qquad \qquad \text{R}' \\ \qquad \qquad \\ \text{HOO}-\text{C}-\text{(OO}-\text{C}-\text{)}_x-\text{OOH} \\ \qquad \qquad \\ \text{R} \qquad \qquad \text{R} \end{array}$ + other structures	not applicable

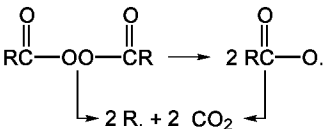
^a $x = 0$ or 1.
^b Temperature at which $t_{1/2} = 10$ h. In benzene.
^c In trichloroethylene (TCE).

Peroxide half-life data provide useful guidance for comparing the activity of one peroxide with another in a given application, if the previously discussed limitations of half-life data are considered. One producer of organic peroxides provides customers with extensive half-life data on its commercial and developmental organic peroxides (21,29). In addition, customer guidance is provided for selection of organic peroxides for various commercial applications, eg, vinyl monomer polymerizations, curing of unsaturated polyester resins, cross-linking of elastomers and polyolefins, and reactive extrusion, based on peroxide type and half-life criteria. This information is provided in a half-life package consisting of a half-life bulletin and associated personal computer interactive software. With the provided software, the customer can rapidly find the optimal time and temperature conditions for utility of an

initiator in a given application. Software is also provided which allows determination of half-life data (rate constants, half-life times, half-life temperatures, activation energies, *A* values, enthalpies of activation, and entropies of activation) on other peroxides, aliphatic azo compounds, carbon–carbon initiators, etc, based on rate–temperature data provided in the chemical literature (20).

Table 2 shows that commercial organic peroxides are available with 10-h half-life temperature activity varying from about room temperature to about 130°C. Organic peroxide classes such as diacyl peroxides and peroxyesters show a strong correlation between structural variation and 10-h half-life temperature activity. Other organic peroxide classes, eg, peroxydicarbonates and monoperoxycarbonates, show very little change in activity with structural variation. The diperoxyketals and dialkyl peroxides show a moderate change in activity with variation in peroxide structures. In the cases of hydroperoxides and ketone peroxides, precise half-life data are difficult to obtain owing to the susceptibilities of these thermally stable peroxide classes to induced decompositions and transition-metal catalysis. Furthermore, radicals are usually generated from these two classes of peroxides at lower temperatures using activators (or promoters), and first-order decomposition rates have no significance. Although the low temperature acyl sulfonyl peroxide, acetyl cyclohexanesulfonyl peroxide (ACSP) [3179-56-4] (with a 10-h half-life temperature of 42°C), is still used to some extent in bulk vinyl chloride polymerizations (30,31), it is only produced captively; hence its peroxide class was not included in Table 2.

Diacyl Peroxides. Table 3 lists several commercial diacyl peroxides and their corresponding 10-h half-life temperatures, determined in benzene and other solvents (32). Although diacyl peroxides cleave at the oxygen–oxygen bond, decarboxylation can occur, either simultaneously or subsequently (eq. 22):



The extent of decarboxylation primarily depends on temperature, pressure, and the stability of the incipient R· radical. The more stable the R· radical, the faster and more extensive the decarboxylation. With many diacyl peroxides, decarboxylation and oxygen–oxygen bond scission occur simultaneously in the transition state. Acyloxy radicals are known to form initially only from diacetyl peroxide and from dibenzoyl peroxides (because of the relative instabilities of the corresponding methyl and phenyl radicals formed upon decarboxylation). Diacyl peroxides derived from non- α -branched carboxylic acids, eg, dilauroyl peroxide, may also initially form acyloxy radical pairs; however, these acyloxy radicals decarboxylate very rapidly and the initiating radicals are expected to be alkyl radicals. Diacyl peroxides are also susceptible to induced decompositions:

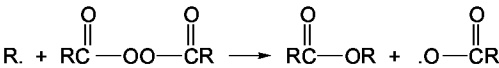
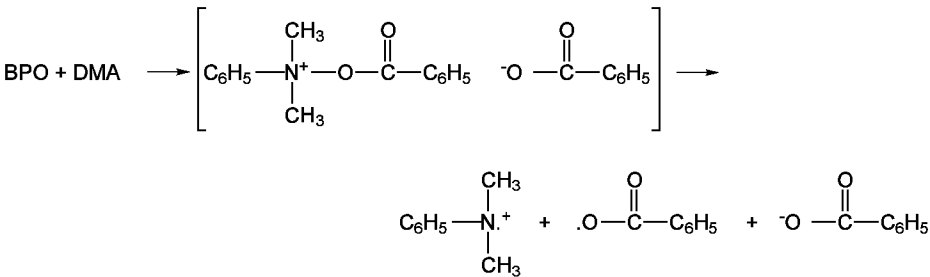


Table 3. Commercial Diacyl Peroxides

Name	CAS Registry Number	10-h $t_{1/2}$, °C ^a	Solvent
dibenzoyl peroxide	[94-36-0]	73	benzene
$\text{C}_6\text{H}_5-\overset{\text{O}}{\parallel}{\text{C}}-\text{OO}-\overset{\text{O}}{\parallel}{\text{C}}-\text{C}_6\text{H}_5$ R in			
diacetyl peroxide	[110-22-5]	72	TCE
		69	benzene
$\text{CH}_3-\overset{\text{O}}{\parallel}{\text{C}}-\text{OO}-\overset{\text{O}}{\parallel}{\text{C}}-\text{CH}_3$ R in			
dilauroyl peroxide	[105-74-8]	62	benzene
$\text{CH}_3(\text{CH}_2)_{10}-\overset{\text{O}}{\parallel}{\text{C}}-\text{OO}-\overset{\text{O}}{\parallel}{\text{C}}-\text{CH}_3(\text{CH}_2)_{10}$ R in			
succinic acid peroxide	[123-23-9]	64	TCE
		66	acetone
$\text{HOOCCH}_2\text{CH}_2-\overset{\text{O}}{\parallel}{\text{C}}-\text{OO}-\overset{\text{O}}{\parallel}{\text{C}}-\text{HOOCCH}_2\text{CH}_2$ R in			
diisononanoyl peroxide	[58499-37-9]	61	TCE
$\text{C}_8\text{H}_{17}-\overset{\text{O}}{\parallel}{\text{C}}-\text{OO}-\overset{\text{O}}{\parallel}{\text{C}}-\text{C}_8\text{H}_{17}$ R in			

^a Temperature at which $t_{1/2}$ = 10 h.

Diacyl peroxides are used in a broad spectrum of applications, including curing of unsaturated polyester resin compositions, cross-linking of elastomers, production of poly(vinyl chloride), polystyrene, and polyacrylates, and in many nonpolymeric addition reactions. Aromatic diacyl peroxides such as dibenzoyl peroxide (BPO) [94-36-0] may be used with promoters to lower the useful decomposition temperatures of the peroxides, although usually with some sacrifice to radical generation efficiency. The most widely used promoter is dimethylaniline (DMA). The BPO–DMA combination is used for hardening (curing) of unsaturated polyester resin compositions, eg, body putty in auto repair kits. Here, the aromatic *tert*-amine promoter attacks the BPO to initially form *N*-benzoyloxydimethylanilinium benzoate (ion pair) which subsequently decomposes at room temperature to form a benzoate ion, a dimethylaniline radical cation, and a benzoyloxy radical that, in turn, initiates the curing reaction (33):



Whereas the BPO–DMA redox system works well for curing of unsaturated polyester blends, it is not a very effective system for initiating vinyl monomer polymerizations, and therefore it generally is not used in such applications (34). However, combinations of amines (eg, DMA) and acyl sulfonyl peroxides (eg, ACS_P) are very effective initiator systems at 0°C for high conversion suspension polymerizations of vinyl chloride (35). BPO has also been used in combination with ferrous ammonium sulfate to initiate emulsion polymerizations of vinyl monomers via a redox reaction (36).

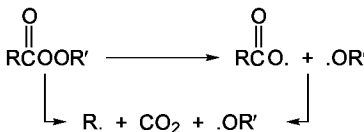
tert-Alkyl Peroxyesters. Table 4 lists several commercial *tert*-alkyl peroxyesters and their corresponding 10-h half-life temperatures (determined in dodecane and other solvents) (37). Only *tert*-alkyl peroxyesters are commercially available. As illustrated in Table 2, the peroxyester class offers the broadest range of temperature activity of any of the peroxide classes.

Table 4. Commercial *tert*-Alkyl Peroxyesters

Name	CAS Registry Number	Structure	10-h $t_{1/2}^{a,b}$, °C
<i>tert</i> -butyl peroxybenzoate	[614-45-9]		104
<i>tert</i> -butyl peroxyacetate	[107-71-1]		102
<i>tert</i> -butyl 2-ethylperoxy-hexanoate ^c	[3006-82-4]		77
<i>tert</i> -butyl peroxy-pivalate	[927-07-1]		62
2,5-di(2-ethylhexa- <i>n</i> oylperoxy)-2,5-dimethyl-hexane	[13052-09-0]		73
<i>tert</i> -butyl peroxy-maleate	[1931-62-0]		87
<i>tert</i> -amyl-2-ethylperoxy-hexanoate	[686-31-7]		75
α-cumyl peroxyneoheptanoate	[104852-44-0]		43
3-hydroxy-1,1-dimethyl-butyl peroxy-neoheptanoate	[110972-57-1]		41

^a Temperature at which $t_{1/2} = 10$ h.
^b In dodecane, unless otherwise noted.
^c *tert*-Butyl peroxyoctate.

Peroxyesters undergo single- or multiple-bond scission to generate acyloxy and alkoxy radicals, or alkyl and alkoxy radicals and carbon dioxide:



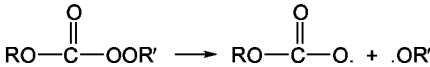
Acyloxy radicals can decarboxylate, as noted above for the diacyl peroxides. The alkoxy radicals (R'O·) can undergo the β-scission reaction leading to greater radical reaction selectivity. Variation of the R group or the R' group provides a convenient means of altering the relative activity of peroxyesters. For example, increasing the steric bulk of either or both of R and R' generally lowers the thermal stability of a peroxyester. Thermal stability decreases as follows: for R, CH₃ > RCH₂ > R₂CH > R₃C. For R', *tert*-butyl > *tert*-amyl > *tert*-octyl > α-cumyl > 3-hydroxy-1,1-dimethylbutyl.

By way of example, *tert*-butyl peroxyacetate [107-71-1] is more thermally stable than 3-hydroxy-1,1-dimethylbutyl peroxyneohexanoate [110972-57-1]. Although other factors affect thermal stability, the trends shown can be used to qualitatively predict peroxyester reactivity trends. The order of activity of the R' group in peroxyesters is also observed in other *tert*-alkylperoxy-containing compounds.

Peroxyesters, particularly those with α-hydrogens or conjugated double bonds, are susceptible to induced decomposition under certain conditions, but they are generally less susceptible than diacyl peroxides. Lower molecular weight peroxyesters that have some water solubility can be hydrolyzed.

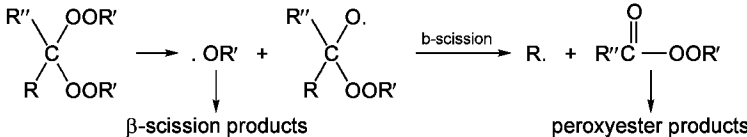
The more selective nature of the radicals produced by *tert*-amyl peroxyesters and other *tert*-amyl peroxides has led to their use in commercial polymer applications requiring discriminating radicals, such as polyol grafting and high solids acrylic resin production (38). *tert*-Amyl peroxides have been replacing aliphatic cyanoazo initiators in these applications. Owing to their diverse structures and associated reactivities, peroxyesters are also used in many other applications, including polymerization of ethylene, vinyl chloride, styrene and acrylate esters, and curing of unsaturated polyester resins.

Monoperoxycarbonates. *OO-tert*-Alkyl *O*-alkyl monoperoxycarbonates (37) (eg, *OO-tert*-butyl *O*-isopropyl monoperoxycarbonate [2372-21-6]) are a class of peroxides related to peroxyesters that also generate alkoxy radicals, ·OR', which again as above can undergo β-scission.



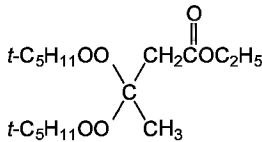
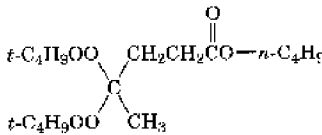
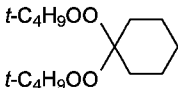
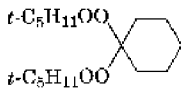
Changing the structure of R' affects the activity of monoperoxycarbonates as previously discussed for peroxyesters. The other cogenerated radical is an alkoxy-carbonyloxy radical. The nature of the R group has practically no effect on the reactivity of monoperoxycarbonates having the same *OO-tert*-alkyl group. The 10-h half-life temperature remains at 100°C for almost all *OO-tert*-butyl *O*-alkyl monoperoxycarbonates.

Diperoxyketals. Some commercially available di(*tert*-alkylperoxy)ketals and their corresponding 10-h half-life temperatures (determined in dodecane) are listed in Table 5 (39). Diperoxyketals thermally decompose by cleavage of only one oxygen–oxygen bond initially, usually followed by β-scission of the resulting alkoxy radicals (40). For acyclic diperoxyketals, β-scission produces an alkyl radical and a peroxyester.



Owing to similarity of thermal stability, the peroxyester decomposes, as discussed previously. Cyclic diperoxyketals such as 1,1-di(*tert*-butylperoxy)cyclohexane cleave the cycloalkyl ring during β-scission to give an alkyl radical with an attached peroxyester group. The effect, after peroxyester decomposition, is the production of two monoradicals (·OR'), a diradical (·R'–R·), and carbon dioxide. Because of the generation of diradicals, cyclic diperoxyketals such as 1,1-di(*tert*-butylperoxy)cyclohexane are effective in enhancing polymer molecular weight or increasing polymer productivity when employed as initiators in commercial vinyl monomer polymerizations (41,42). Diperoxyketals are used commercially in styrene polymerizations, curing of elastomers, and in elevated temperature curing of unsaturated polyester resin compositions. *tert*-Amyl diperoxyketals are good initiators for acrylics, especially in the preparation of high solids coatings resins (43).

Table 5. Commercial Diperoxyketals^a

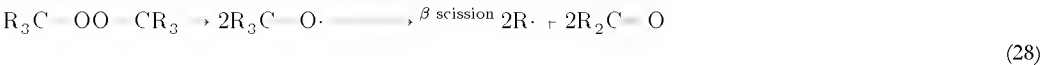
Name	CAS Registry Number	Structure	10-h t _{1/2} ^{b,c} , °C
ethyl 3,3-di-(<i>tert</i> -amylperoxy)-butyrate	[67567-23-1]		112
<i>n</i> -butyl 4,4-di(<i>tert</i> -butylperoxy)-valerate	[995-33-5]		109
1,1-di(<i>tert</i> -butyl-peroxy)cyclohexane	[3006-86-8]		97
1,1-di(<i>tert</i> -amyl-peroxy)cyclohexane	[15667-10-4]		93

^a Ref. 39.

^b Temperature at which t_{1/2} = 10 h.

^c In dodecane.

Dialkyl Peroxides. Some commercially available dialkyl peroxides and their corresponding 10-h half-life temperatures in dodecane are listed in Table 6 (44). Dialkyl peroxides initially cleave at the oxygen–oxygen bond to generate alkoxy radical pairs:



Because high temperatures are required to decompose dialkyl peroxides at useful rates, β-scission of the resulting alkoxy radicals is more rapid and more extensive than for most other peroxide types. When methyl radicals are produced from alkoxy radicals, the dialkyl peroxide precursors are very good initiators for cross-linking, grafting, and degradation reactions. When higher alkyl radicals such as ethyl radicals are produced, the dialkyl peroxides are useful in vinyl monomer polymerizations.

Table 6. Commercial Dialkyl Peroxides^a

Name	CAS Registry Number	Structure	10-h <i>t</i> _{1/2} ^{b,c} , °C
2,5-di(<i>tert</i> -butylperoxy)-2,5-dimethyl-3-hexyne	[1068-27-5]		131
2,5-di(<i>tert</i> -butylperoxy)-2,5-dimethylhexane	[78-63-7]		120
di- <i>tert</i> -butyl peroxide	[110-05-4]	<i>t</i> -C ₄ H ₉ -OO- <i>t</i> -C ₄ H ₉	129
di- <i>tert</i> -amyl peroxide	[10508-09-5]	<i>t</i> -C ₅ H ₁₁ -OO- <i>t</i> -C ₅ H ₁₁	123
dicumyl peroxide	[80-43-3]		117

^a Ref. 44.
^b Temperature at which *t*_{1/2} = 10 h.
^c In dodecane.

Dialkyl Peroxydicarbonates. Some commercially available dialkyl peroxydicarbonates and their corresponding 10-h half-life temperatures (determined in trichloroethylene solutions) are listed in Table 7 (45). These peroxides are active at low temperatures and initially undergo homolytic cleavage to produce alkoxy-carbonyloxy radical pairs that may subsequently decarboxylate to produce alkoxy radicals:

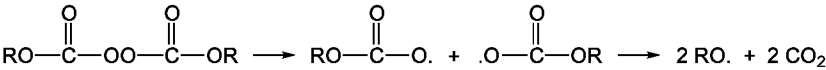


Table 7 shows that the nature of the alkyl group, whether primary alkyl, secondary alkyl, or cycloalkyl, does not affect the 10-h half-life temperatures of dialkyl peroxydicarbonates in trichloroethylene (TCE). All peroxydicarbonates have about the same 10-h half-life temperature in TCE (49–50°C).

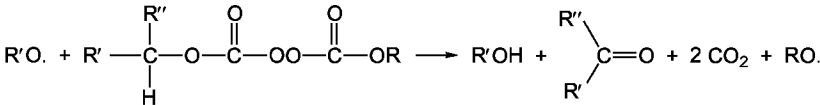
Table 7. Some Commercial Dialkyl Peroxydicarbonates^a

Name	CAS Registry Number	Structure	10-h <i>t</i> _{1/2} ^{b,c} , °C
di- <i>n</i> -propyl peroxydicarbonate	[16066-38-9]		50
di- <i>sec</i> -butyl peroxydicarbonate	[19910-65-7]		50
di-(2-ethylhexyl)-peroxydicarbonate	[16111-62-9]		49
dicyclohexyl peroxydicarbonate	[1561-49-5]		50

^a Ref. 45.
^b Temperature at which *t*_{1/2} = 10 h.

^c In TCE.

As a peroxide class, dialkyl peroxydicarbonates are very susceptible to radical-induced decompositions:



Decomposition rate studies on dialkyl peroxydicarbonates in various solvents reveal dramatic solvent effects that primarily result from the susceptibility of peroxydicarbonates to induced decompositions. These studies show a decreasing order of stability of peroxydicarbonates in solvents as follows: TCE > saturated hydrocarbons > aromatic hydrocarbons > ketones (29). Decomposition rates are lowest in TCE where radicals are scavenged before they can induce the decomposition of peroxydicarbonate molecules.

Peroxydicarbonates are efficient polymerization initiators for most vinyl monomer polymerizations, especially for monomers such as acrylates, ethylene, and vinyl chloride. They are particularly good initiators for less reactive monomers such as those containing allyl groups. They are also effective for curing of unsaturated polyester molding resins.

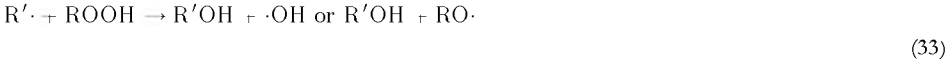
tert-Alkyl Hydroperoxides. Some commercially available *tert*-alkyl hydroperoxides (46) are listed in Table 8. Hydroperoxides can decompose thermally to initially form alkoxy and hydroxy radicals:



However, because of the high temperature nature of this class of peroxides (10-h half-life temperatures of 133–172°C) and their extreme sensitivities to radical-induced decompositions and transition-metal activation, hydroperoxides have very limited utility as thermal initiators. The oxygen–hydrogen bond in hydroperoxides is weak (368–377 kJ mol (88.0–90.1 kcal mol) BDE) and is susceptible to attack by higher energy radicals:

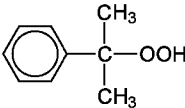
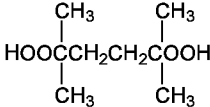
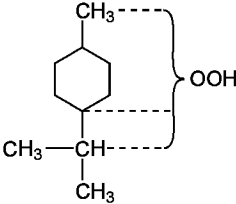
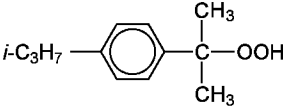


Further reactions of the alkylperoxy radical (ROO·) depend on the environment but generally cause generation of other radicals that can attack undecomposed hydrosend peroxide, thus perpetuating the induced decomposition chain. Radicals also can attack undecomposed peroxide by radical displacement on the oxygen–oxygen bond:



This is basically the same type of induced decomposition that occurs with other peroxide classes, eg, the dialkyl peroxydicarbonates and diacyl peroxides.

Table 8. Commercial *tert*-Alkyl Hydroperoxides^a

Name	CAS Registry Number	Structure
<i>tert</i> -butyl hydroperoxide	[75-91-2]	<i>t</i> -C ₄ H ₉ OOH
<i>tert</i> -amyl hydroperoxide	[3425-61-4]	<i>t</i> -C ₅ H ₁₁ OOH
α-cumyl hydroperoxide	[80-15-9]	
2,5-dihydroperoxy-2,5-dimethylhexane	[3025-88-5]	
<i>para</i> -menthane hydroperoxide ^b	[26762-92-5]	
<i>m p</i> -isopropyl-α-cumyl hydroperoxide ^c	[98-49-7]	

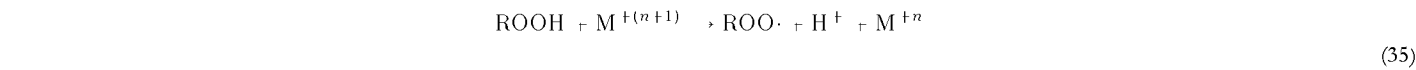
^a Ref. 46.

^b OOH may be attached to any of the three positions indicated.

^c Para isomer shown.

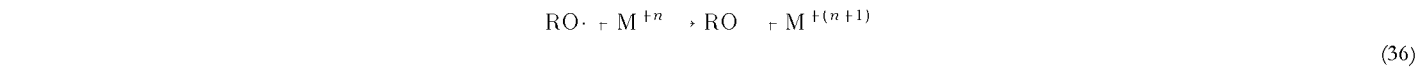
Hydroperoxides are more widely used as initiators in low temperature applications (at or below room temperature) where transition-metal (M) salts are employed as activators. The activation reaction involves electron-transfer (redox) mechanisms:





Either oxidation state of a transition metal (Fe, Mn, V, Cu, Co, etc) can activate decomposition of the hydroperoxide. Thus a small amount of transition-metal ion can decompose a large amount of hydroperoxide. Trace transition-metal contamination of hydroperoxides is known to cause violent decompositions. Because of this fact, transition-metal promoters should never be premixed with the hydroperoxide. Trace contamination of hydroperoxides (and ketone peroxides) with transition metals or their salts must be avoided.

Transition-metal ions also react with the generated radicals to convert the radicals to ions:



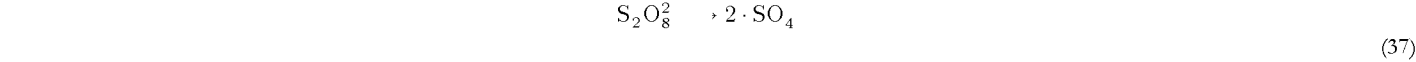
This reaction is one example of several possible radical transition-metal ion interactions. The significance of this and similar reactions is that radicals are destroyed and are no longer available for initiation of useful radical reactions. Consequently, the optimum use levels of transition metals are very low. Although the hydroperoxide decomposes quickly when excess transition metal is employed, the efficiency of radical generation is poor.

Ketone Peroxides. These materials are mixtures of compounds with hydroperoxy groups and are composed primarily of the two structures shown in Table 2. Ketone peroxides are marketed as solutions in inert solvents such as dimethyl phthalate. They are primarily employed in room-temperature-initiated curing of unsaturated polyester resin compositions (usually containing styrene monomer) using transition-metal promoters such as cobalt naphthenate. Ketone peroxides contain the hydroperoxy (—OOH) group and thus are susceptible to the same hazards as hydroperoxides. By far the most popular commercial ketone peroxide is methyl ethyl ketone peroxide [1338-23-4]. Smaller quantities of ketone peroxides such as methyl isobutyl ketone peroxide [28056-59-9], cyclohexanone peroxide [12262-58-7], and 2,4-pentanedione peroxide [37187-22-7] are used commercially (47).

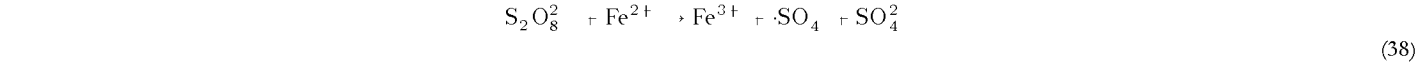
Selection of organic peroxides for various commercial applications has been reviewed (21,33,48), particularly for vinyl chloride polymerizations (30).

Inorganic Peroxides

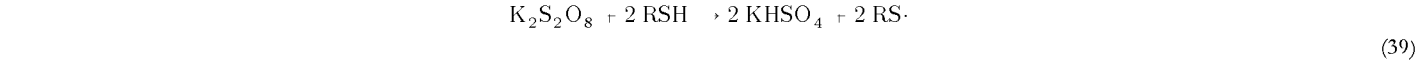
Inorganic peroxide–redox systems have been employed for initiating emulsion homo- and copolymerizations of vinyl monomers. These systems include hydrogen peroxide–ferrous sulfate, hydrogen peroxide–dodecyl mercaptan, potassium peroxydisulfate–sodium bisulfite, and potassium peroxydisulfate–dodecyl mercaptan (36,49). Potassium peroxydisulfate, K₂S₂O₈ [7727-21-2] (or the corresponding sodium or ammonium salt), is an inorganic peroxide that is used widely in emulsion polymerization (eg, latexes, rubbers, etc), usually in combination with a reducing agent. Without reducing agents, the peroxydisulfate ion decomposes to give sulfate ion radicals:



With transition-metal activators, the initiation process is postulated as:



The reaction with mercaptans is believed to generate initiating sulfur radicals:

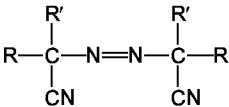


Hydrogen peroxide, in combination with reducing agents (transition metals), also is used in those applications where its high water- and low oil-solubility is not a problem or is easily overcome.

When handling and using peroxide initiators, care should be exercised since they are thermally sensitive and decompose (sometimes violently) when exposed to excessive temperatures, especially when they are in their pure or highly concentrated states. However, they are useful as initiators because of their thermal instability. What may be a safe temperature for one peroxide can be an unsafe temperature for another, since peroxide initiators encompass a wide activity range. Because some peroxides are shock- or friction-sensitive in the pure state, they are generally desensitized by formulating them into solutions, pastes, or powders with inert diluents. All manufacturers' literature should be carefully scrutinized and the peroxide safety literature should be reviewed before handling and using specific peroxide initiator compositions (31,32,37,39,44–48,50–53).

Azo Compounds

Generally, the commercially available azo initiators are of the symmetrical azonitrile type:

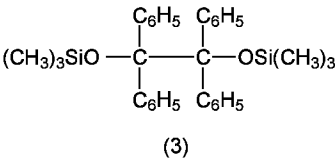
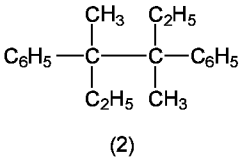
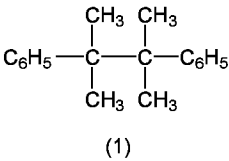


The symmetrical azonitriles are solids with limited solubilities in common solvents (54–56). Some commercial aliphatic azo compounds and their 10-h half-life temperatures are listed in Table 9.

Table 9. Commercial Azo Initiators

Name	CAS Registry Number	Structure	10-h <i>t</i> _{1/2} ^a , °C	Solvent
2,2'-azobis[4-methoxy-2,4-dimethyl]pentanenitrile	[15545-97-8]	$\begin{array}{ccccccc} & \text{OCH}_3 & & \text{CH}_3 & & \text{CH}_3 & \text{OCH}_3 \\ & & & & & & \\ \text{CH}_3 - & \text{C} - \text{CH}_2 - & \text{C} - \text{N} = \text{N} - & \text{C} - \text{CH}_2 - & \text{C} - \text{CH}_3 \\ & & & & & & \\ & \text{CH}_3 & & \text{C} \equiv \text{N} & & \text{C} \equiv \text{N} & \text{CH}_3 \end{array}$	33	toluene
2,2'-azobis[2,4-dimethyl]pentanenitrile	[4419-11-8]	$\begin{array}{ccccccc} & \text{CH}_3 & & \text{CH}_3 & & \text{CH}_3 & \text{CH}_3 \\ & & & & & & \\ \text{HC} - & \text{CH}_2 - & \text{C} - \text{N} = \text{N} - & \text{C} - \text{CH}_2 - & \text{CH} \\ & & & & & & \\ & \text{CH}_3 & & \text{C} \equiv \text{N} & & \text{C} \equiv \text{N} & \text{CH}_3 \end{array}$	52	toluene
2,2'-azobis[isobutyronitrile]	[78-67-1]		64	toluene

Three carbon–carbon initiators are currently available commercially, 2,3-dimethyl-2,3-diphenylbutane [1889-67-4] (1), 3,4-dimethyl-3,4-diphenylhexane [10192-93-5] (2), and 1,1,2,2-tetraphenyl-1,2-bis(trimethylsiloxy)ethane [22341-08-8] (3).



Initiators (1) and (2) have 10-h half-life temperatures of 237°C and 201°C, respectively. It has been reported that, unlike organic peroxides and aliphatic azo compounds, carbon–carbon initiators (1) and (2) undergo endothermic decompositions (62). These carbon–carbon initiators are useful commercially as fire-retardant synergists in fire-resistant expandable polystyrenes (63).

Other Radical Generating Systems

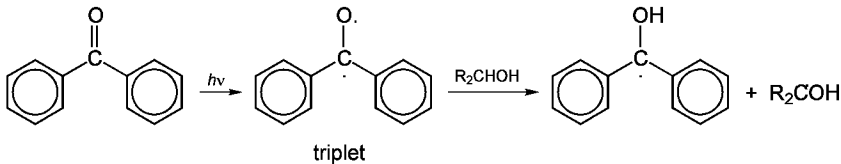
There are many chemical methods for generating radicals reported in the literature that do not involve conventional initiators. Specific examples are included in References 64–79. Most of these radical-generating systems cannot broadly compete with the use of conventional initiators in industrial polymer applications owing to cost or efficiency considerations. However, some systems may be well-suited for initiating specific radical reactions or polymerizations, eg, grafting of monomers to cellulose using ceric ion (80).

Initiation Through Radiation and Photoinitiators

High energy ionizing radiation sources (eg, x-rays, γ-rays, α-particles, β-particles, fast neutrons, and accelerator-generated electrons) can generate radical sites on organic substrates (81). If the substrate is a vinyl monomer, radical polymerization can occur (82). If the substrate consists of a polymer and a vinyl monomer, then polymer cross-linking, degradation, grafting of the monomer to the polymer, and polymerization of the monomer can all occur (83). Radical polymerizations of vinyl monomers with ionized plasma gases have been reviewed (84).

Initiation of radical reactions with uv radiation is widely used in industrial processes (85). In contrast to high energy radiation processes where the energy of the radiation alone is sufficient to initiate reactions, initiation by uv irradiation usually requires the presence of a photoinitiator, ie, a chemical compound or compounds that generate initiating radicals when subjected to uv radiation. There are two types of photoinitiator systems: those that produce initiator radicals by intermolecular hydrogen abstraction and those that produce initiator radicals by photocleavage (86–91).

In the case of intermolecular hydrogen abstraction, a hydrogen (H) atom donor is required. Typical donors have an active H atom positioned alpha to an oxygen or nitrogen, eg, alcohols (R₂CHOH), ethers (R₂CHOR), and *tert*-amines (R₂CHNR₂), or an active H atom directly attached to sulfur, eg, thiols (RSH). Some of the commercial photoinitiators that undergo intermolecular H abstraction from the H atom donor upon excitation by uv radiation are listed in Table 10. A reaction illustrating this photoinitiation process is given below for benzophenone (photoinitiator) and an alcohol (H atom donor):



Upon exposure to uv light, ground-state benzophenone is excited to the triplet state (a diradical) which abstracts an alpha H atom from the alcohol, resulting in the formation of two separate initiating radicals. With amine H atom donors, an electron transfer may precede the H-transfer, as in triplet exciplex formation between benzophenone and amine (eq. 43):

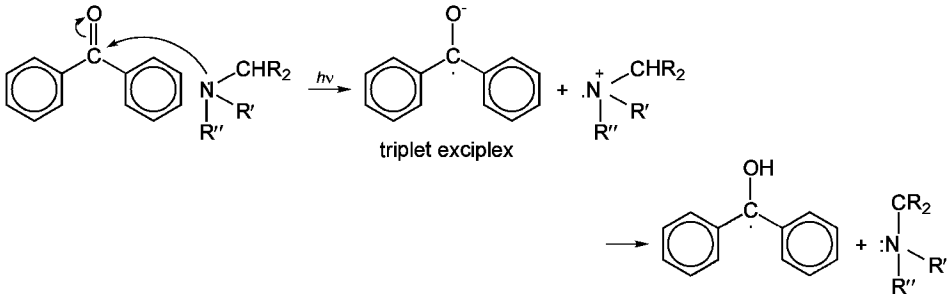
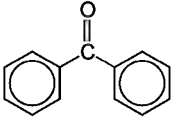
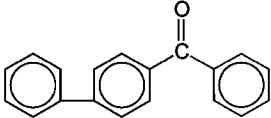
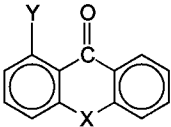
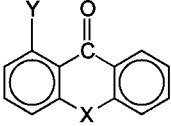
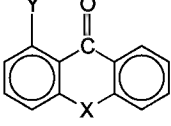
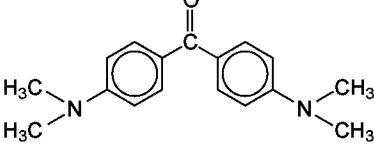
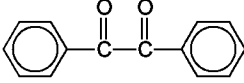
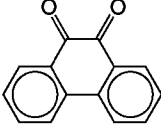
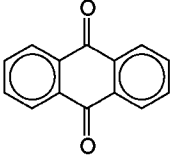


Table 10. Photoinitiators that Undergo Intermolecular H Abstraction

Name	CAS Registry Number	Structure
benzophenone	[119-61-9]	

		
4-phenylbenzophenone	[2128-93-0]	
		
xanthone (X = O, Y = H)	[90-47-1]	
		
thioxanthone (X = S, Y = H)	[492-22-8]	
		
2-chlorothioxanthone (X = S, Y = Cl)	[86-39-5]	
		
4,4'-bis(N,N'-dimethylamino)benzophenone (Michler's ketone)	[90-94-8]	
		
benzil	[134-81-6]	
		
9,10-phenanthraquinone	[84-11-7]	
		
9,10-anthraquinone	[84-65-1]	
		

Some commercial photoinitiators (Table 11) undergo a Norrish Type I photocleavage to form two initiating radical fragments directly for a benzoin ether:

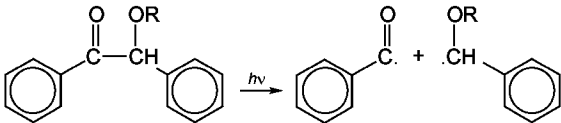
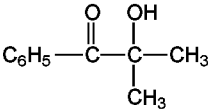
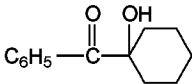


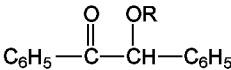
Table 11. Photoinitiators That Undergo Photocleavage

Name	CAS Registry Number	Structure
α,α -dimethyl- α -hydroxy-acetophenone	[7473-98-5]	

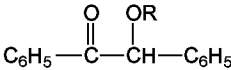
(1-hydroxycyclohexyl)-phenylmethanone [947-19-3]



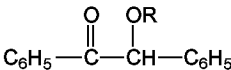
benzoin ethers
benzoin methyl ether [3524-62-7]



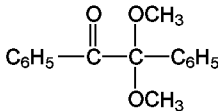
benzoin ethyl ether [574-09-4]



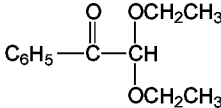
benzoin isobutyl ether [22499-12-3]



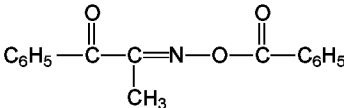
α,α -dimethoxy- α -phenyl-acetophenone [24650-42-8]



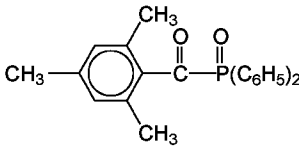
α,α -diethoxyacetophenone [6175-45-7]



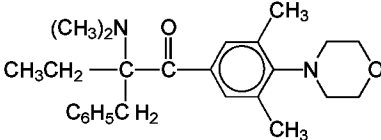
1-phenyl-1,2-propanedione,
2-(O-benzoyl)oxime [17292-57-8]



diphenyl(2,4,6-trimethyl-benzoyl)phosphine
oxide [75980-60-8]



α -dimethylamino- α -ethyl- α -benzyl-3,5-dimeth
yl-4-morpholino-acetophenone [119313-12-1]



In many photoinitiated processes, a photosensitizer may be used. A photosensitizer absorbs light and subsequently transfers the absorbed energy to an energy acceptor, which then can produce initiator radicals by H abstraction or by photocleavage. The energy-transfer agent (photosensitizer) usually undergoes no net change. A variety of photosensitizers have been used such as eosin, chlorophyll, methylene blue, and thioxanthone. In photosensitized processes, the energy acceptor often is referred to as a co-initiator. These co-initiators do not absorb light but accept energy from the excited photosensitizer, which distinguishes them from the photoinitiators listed in Tables 10 and 11. Typical co-initiators that undergo H abstraction are the H donors mentioned above. An example of a co-initiator undergoing photocleavage is quinoline-8-sulfonyl chloride [18704-37-5] photosensitized by thioxanthone (86).

The peroxide and azo thermal initiators also are photochemically unstable and have been used as radical sources at well below their normal thermal decomposition temperatures. However, their industrial use as photoinitiators has been limited because their light-absorption characteristics frequently are unsuitable and because of the obvious potential complication owing to their slow thermal decomposition, which leads to poor shelf-life and nonreproducible photoactivity in given formulations (88). Further information on photoinitiators can be found in the literature (92).

Economic Aspects

The principal worldwide producers of organic peroxide initiators (and their trade names) include Elf Atochem (Luperco, Luperox, Lupersol, Lucidol (US), Luchem, Alperox, Decanox, Peroximon, and Retilox), Akzo (Trigonox, Perkadox, Cadox, Cadet, Laurox, Liladox, Kenodox, Lucidol (Europe), Butanox, Cyclonox), Aztec (Aztec), Peroxid-Chemie (Interox), Witco (Esperox, Esperal, USP, Quickset, Hi Point), Nippon Oil & Fats Co. (Nyper, Perbutyl, Percumyl, Perthexa, Permek, Peroyl), Norac (Superox), Hercules (DiCup, VulCup), and Sanken Kako (Sanperox). Worldwide the three leading producers of organic peroxides are Akzo, Elf Atochem, and Peroxid-Chemie. Approximate U.S. sales of organic peroxide initiators in 1992 were ~38,640,000 kg, valued at \$200 × 10⁶; worldwide sales were 109 × 10⁶ kg valued at \$750 × 10⁶. The principal worldwide producers of organic azo initiators are Du Pont (Vazo), Elf Atochem, and Wako. The worldwide market for organic azo initiators is small, being only about 10⁰ % of the market for organic peroxide initiators. Ciba Geigy is a significant supplier of photoinitiators (Darocur, Irgacure). Sales figures on photoinitiators are not readily available. The market for these initiators has been reviewed (93). Because most of the consumption of organic peroxides and azo initiators is in the developed countries, market growth in the 1990s is expected to be modest, ie, 2–3^o annually.

BIBLIOGRAPHY

"Initiators" in *ECT* 3rd ed., Vol. 13, pp. 355–373, by C. S. Sheppard and V. Kamath, Pennwalt Corp.

1. D. Griller and K. U. Ingold, *Acc. Chem. Res.* **9**(1), 13 (1976).
2. C. H. Bamford, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, 2nd ed., Vol. 13, Wiley-Interscience, New York, 1988, pp. 708–867.
3. W. A. Pryor, ed., *Organic Free Radicals*, *ACS Symposium Series 69*, American Chemical Society, Washington, D.C., 1978; J. M. Hay, *Reactive Free Radicals*, Academic Press, Inc., New York, 1974.
4. J. K. Kochi, ed., *Free Radicals*, Vols. I and II, John Wiley & Sons, Inc., New York, 1973.
5. C. Walling, *Free Radicals in Solution*, John Wiley & Sons, Inc., New York, 1957.
6. W. G. Lloyd, *Chemtech.*, 176 (Mar. 1971); 371 (June 1971); 687 (Nov. 1971); 182 (Mar. 1972).
7. J. M. Tedder and J. C. Walton, *Acc. Chem. Res.* **9**, 183 (1976).
8. W. A. Pryor, *Chem. Eng. News*, 70 (Jan. 15, 1968).
9. I. P. Gragerov, *Russ. Chem. Rev.* **38**, 626 (1969).
10. K. U. Ingold, *Free Radical Substitution Reactions*, Wiley-Interscience, New York, 1971.
11. E. S. Huyser, *Free Radical Chain Reactions*, Wiley-Interscience, New York, 1970.
12. G. Sosnovsky, *Free Radical Reactions in Preparative Organic Chemistry*, Macmillan, New York, 1964.
13. S. Patai, ed., *The Chemistry of Peroxides*, John Wiley & Sons, Inc., New York, 1983; S. Patai, ed., *Supplement E2: The Chemistry of Hydroxyl, Ether and Peroxide Groups*, John Wiley & Sons, Inc., New York, 1993.
14. S. Patai, ed., *The Chemistry of the Hydrazo, Azo, and Azoxy Groups*, John Wiley & Sons, Inc., New York, 1975.
15. W. Ando, ed., *Organic Peroxides*, John Wiley & Sons, Inc., New York, 1992.
16. A. A. Zavitsas, *J. Am. Chem. Soc.* **94**, 2779–2788 (1972); R. T. Sanderson, *J. Am. Chem. Soc.* **97**, 1367 (1975); D. M. Golden and S. W. Benson, *Chem. Rev.* **69**, 125 (1969); P. Gray, A. A. Herod, and A. Jones, *Chem. Rev.* **71**, 247 (1971).
17. G. Leroy, *Mol. Struct.* **120**, 91–106 (1985); J. M. Tedder, *Tetrahedron* **38**, 313 (1982).
18. J. M. Tedder and J. C. Walton, *Tetrahedron* **36**, 701 (1980).
19. D. Lefort, *New J. Chem.* **16**, 219–232 (1992); C. Ruchardt, *Free Rad. Res. Comms.* **2**(4–6), 197 (1987).
20. J. C. Masson, in J. Brandrup and E. H. Immergut, eds., *Polymer Handbook*, 3rd ed., John Wiley & Sons, Inc., New York, 1989.
21. *Technical Publication, HALFLIFE—Peroxide Selection Based on Half-Life*, 2nd ed., and Copyrighted Software, Elf Atochem North America, Inc., Philadelphia, Pa., 1992.
22. A. G. Davies, *Organic Peroxides*, Butterworths, London, 1961.
23. E. G. E. Hawkins, *Organic Peroxides*, E. and F. F. Spon Ltd., London, 1961.
24. A. V. Tobolsky and R. B. Mesrobian, *Organic Peroxides*, Interscience Publishers, New York, 1954.
25. D. Swern, ed., *Organic Peroxides*, Vol. I, Wiley-Interscience, New York, 1970.
26. D. Swern, ed., *Organic Peroxides*, Vol. II, Wiley-Interscience, New York, 1971.
27. D. Swern, ed., *Organic Peroxides*, Vol. III, Wiley-Interscience, New York, 1972.
28. *Current Polymerization Catalysts—Peroxide, Azo and Other Initiators*, multi-client study, Catalyst Consultants, Inc., Spring House, Pa., Nov. 1991.
29. P. Callais, V. Kamath, S. Stromberg, J. Sanchez, and J. Gravelle, *Plast. Compound.* **15**(1), 49–52 (1992).
30. *Free Radical Initiators for the Suspension Polymerization of Vinyl Chloride*, *Technical Publication 30.90*, Lucidol Division, Pennwalt Corp., Buffalo, N.Y., 1976.
31. *Organosulfonyl Peroxides*, Product Bulletin, Lucidol Division, Pennwalt Corp., Buffalo, N.Y., 1977.
32. *Dialkyl Peroxides*, Product Bulletin, Organic Peroxide Division, Atochem North America, Inc., Buffalo, N.Y., July 1991.
33. C. S. Sheppard and V. R. Kamath, *Polym. Eng. Sci.* **19**, 597 (1979).
34. Ref. 2, pp. 766–767.
35. **U.S. Pat. 3,629,220** (Dec. 21, 1971), J. Sanchez (to Pennwalt Corporation).
36. H. F. Mark, in N. M. Bikales, ed., *Encyclopedia of Polymer Science and Technology*, Vol. 3, Interscience Publishers, a Division of John Wiley & Sons, Inc., New York, 1965, pp. 26–34.
37. *Peroxyesters*, Product Bulletin, Organic Peroxide Division, Atochem North America, Inc., Buffalo, N.Y., June 1991.
38. **U.S. Pat. 4,777,230** (Oct. 11, 1988), V. R. Kamath (to Pennwalt Corp.).
39. *Peroxyketals*, Product Bulletin, Organic Peroxides, Elf Atochem North America, Inc., Philadelphia, Pa., Oct. 1992.
40. Y. Sugihara, Y. Watanabe, H. Kumura, T. Nakamura, S. Suyama, and Y. Sawaki, *Bull. Chem. Soc. Jpn.* **65**, 664 (1992); Y. Watanabe, K. Ohta, and S. Suyama, *Bull. Chem. Soc. Jpn.* **65**, 2063 (1992); Y. Watanabe, H. Ishigaki, and S. Suyama, *Polym. J.* **24**, 257 (1992); Y. Watanabe, Y. Sugihara, S. Suyama, and T. Nakamura, *Polym. J.* **24**, 971 (1992); K. Matsuyama and H. Kumura, *J. Org. Chem.* **58**, 1766 (1993).
41. V. R. Kamath, *Modern Plast.*, 106, 108, 110 (Sept. 1981).
42. **U.S. Pat. 4,125,695** (Nov. 14, 1978), V. R. Kamath (to Pennwalt Corp.).
43. P. A. Callais, V. R. Kamath, and M. G. Moskal, *Proc. Water-Borne, Higher-Solids, Powder Coat. Symp.* **19**, 156 (1992).
44. *Dialkyl Peroxides*, Product Bulletin, Organic Peroxides, Elf Atochem North America, Inc., Philadelphia, Pa., June 1992.
45. *Peroxydicarbonates*, Product Bulletin, Organic Peroxide Division, Atochem North America, Inc., Buffalo, N.Y., Sept. 1990.
46. *Tertiary Alkyl Hydroperoxides*, Product Bulletin, Atochem North America, Inc., Buffalo, N.Y., Nov. 1991.
47. *Ketone Peroxides*, Product Bulletin, Organic Peroxide Division, Atochem North America, Inc., Buffalo, N.Y., Sept. 1991.
48. C. S. Sheppard, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, 2nd ed., Vol. 11, Wiley-Interscience, New York, 1988, pp. 1–21.
49. D. C. Blackley, *Emulsion Polymerization—Theory and Practice*, John Wiley & Sons, Inc., New York, 1975, Chapt. 6.
50. *Organic Peroxides: Their Safe Handling and Use*, Technical Bulletin 30.40, Organic Peroxides, Elf Atochem North America, Inc., Philadelphia, Pa., Sept. 1992.
51. *Safe Handling, Storage and Transportation of Peroxides Requiring Refrigeration*, Technical Bulletin 30.43, Organic Peroxides, Elf Atochem North America, Inc., Philadelphia, Pa., Sept. 1992.
52. *Suggested Relative Hazard Classification of Organic Peroxides*, Technical Publication, Organic Peroxide Producers Safety Division, The Society of the Plastics Industry, Inc., New York, 1992.
53. *The Storage and Handling of Organic Peroxides in the Reinforced Polyester Fabricating Plant*, Technical Bulletin No. 19, Organic Peroxide Producers Safety Division, The Society of the Plastics Industry, Inc., New York, 1978 revision.
54. Ref. 4, pp. 511–516.
55. *Du Pont Vazo Polymerization Initiators—Properties, Uses, Storage and Handling*, Product Information Bulletin, Du Pont Chemicals, Wilmington, De., July 1984.
56. C. S. Sheppard, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, Vol. 2, 2nd ed., Wiley-Interscience, New York, 1985, pp. 143–157.
57. H. Zollinger, *Azo and Diazo Chemistry, Aliphatic and Aromatic Compounds*, Wiley-Interscience, New York, 1961, Chaps. 9, 12.
58. C. G. Overberger, J.-P. Anselme, and J. G. Lombardino, *Organic Compounds with Nitrogen–Nitrogen Bonds*, The Ronald Press Co., New York, 1966, Chapt. 4.
59. P. A. S. Smith, *The Chemistry of Open-Chain Organic Nitrogen Compounds*, Vol. II, W. A. Benjamin, Inc., New York, 1966, Chapt. 11.

60. K. Ziegler, *Angew. Chem.* **61**, 168 (1949).

61. C. Ruchardt and H. Beckhaus, *Angew. Chem. Int. Ed. Engl.* **19**, 429–440 (1980).

62. *Initiators for Polymer Production—Product Catalog*, Akzo Chemicals, Inc., Chicago, Ill., 1992, pp. 18–19.

63. Technical Information, *Polystyrene: Initiators for Styrene Polymerization*, Akzo Chemie, Deventer, the Netherlands, Nov. 1985.

64. C. H. Bamford, in G. Allen, J. C. Bevington, G. C. Eastmond, A. Ledwith, S. Russo, and P. Sigwalt, eds., *Comprehensive Polymer Science*, Vol. 4, Pergamon Press, Elmsford, N.Y., 1989, pp. 123–139.

65. C. I. Hill and G. M. Whitesides, *J. Am. Chem. Soc.* **96**, 870 (1974).

66. W. Kawai, M. Ogawa, and T. Ichihashi, *J. Polym. Sci. A-1* **9**, 1599 (1971).

67. C. W. Brown and H. M. Longbottom, *J. Appl. Polym. Sci.* **14**, 2927 (1970).

68. J. Bond and P. I. Lee, *J. Polym. Sci. A-1* **6**, 2621 (1968).

69. N. H. Anderson and R. O. C. Norman, *J. Chem. Soc. B*, 993 (1971).

70. M. Imoto, K. Ueda, and K. Takemoto, *Makromol. Chem.* **138**, 11 (1970); J. Barton, V. Durdovic and V. Horanska, *Makromol. Chem.* **133**, 205 (1970).

71. C. M. M. da Silva Correa and W. A. Waters, *J. Chem. Soc. C*, 1880 (1968).

72. T. Otsu and S. Kubota, *Polym. Rep.* **147**, 18 (1970).

73. S. C. Dickerman and co-workers, *J. Org. Chem.* **34**, 714 (1969).

74. M. Kinoshita, N. Yoshizumi, and M. Imoto, *Makromol. Chem.* **127**, 185 (1969).

75. M. Ya. Turkina and I. P. Gragerov, *J. Org. Chem. USSR* **5**, 575 (1969).

76. S. V. Kulkarni, A. E. Fiebig, and R. Filler, *Chem. Ind.* **11**, 364 (1970).

77. E. I. Heiba, R. M. Dessau, and W. J. Koehl, Jr., *J. Am. Chem. Soc.* **91**, 138 (1969).

78. T. Koenig and J. M. Owens, *J. Am. Chem. Soc.* **96**, 4052 (1974); T. Koenig and W. R. Mabey, *J. Am. Chem. Soc.* **92**, 3804 (1970).

79. L. Benati, A. Tundo, and G. Zanardi, *J. Chem. Soc. Chem. Commun.*, 590 (1972); P. Spagnolo and co-workers, *J. Chem. Soc. Perkin I*, 93 (1972).

80. J. C. Arthur, Jr., in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, 2nd ed., Vol. 3, Wiley-Interscience, New York, 1985, pp. 68–86.

81. V. T. Stannett, J. Silverman, and J. L. Gamett, in Ref. 64, pp. 317–336.

82. Ref. 81, pp. 319–323.

83. Ref. 81, pp. 327–329.

84. H. Yasuda and Y. Iriyama, in Ref. 64, pp. 357–375.

85. J. Hutchison, "Photoinitiated Free Radical Chain Reaction," Vol. 11, *Energy Research Abstracts*, Abstract No. 51555, Electricity Council Research Centre, 1986.

86. S. P. Pappas, ed., *UV Curing: Science and Technology*, Technology Marketing Corporation, Stamford, Conn., 1978, Chapt. 1; S. P. Pappas, in Ref. 64, pp. 337–355.

87. H. J. Hageman, *Progr. Org. Coatings* **13**, 123 (1985).

88. A. Ledwith, *Pure Appl. Chem.* **49**, 431 (1977).

89. A. Ledwith, *J. Oil Colour Chem. Assoc.* **59**(5), 157 (1976).

90. A. Pryce, *J. Oil Colour Chem. Assoc.* **59**(5), 166 (1976).

91. B. M. Monroe and G. C. Weed, *Chem. Rev.* **93**, 435–448 (1993).

92. S. P. Pappas, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, 2nd ed., Vol. 11, Wiley-Interscience, New York, 1988, pp. 186–212.

93. C. Verbanic, *Chem. Bus.*, 13–18 (Mar. 1993).

José Sanchez
Terry N. Myers
Elf Atochem North America, Inc.

ANIONIC INITIATORS

Chain-reaction polymerization generally comprises the kinetic steps of initiation, propagation, and termination. Chain-reaction polymerizations can be effected using a variety of initiating species broadly classified as free-radical, cationic, anionic, or coordination (organometallic) initiators, depending on their electrochemical nature. The initiation step generates reactive intermediates (eg, free-radical, anion, cation, or organometallic species) which can add monomer (M) repetitively in a chain reaction as shown in equations 1 and 2 where I is an initiator which generates the initiating species I* in the first step of initiation.



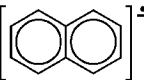
In anionic polymerization, the reactive propagating intermediate generated by the initiation reaction is an anion, ie, a species which carries a formal negative charge, with a corresponding positively charged counterion. In living anionic polymerization, the kinetic steps of chain termination and chain transfer are absent (1–4). This unique aspect of many anionic polymerizations provides a methodology for preparing polymers with control of the significant variables affecting polymer properties including molecular weight, molecular weight distribution, block copolymer composition, and microstructure, as well as molecular architecture (linear, branched, and cyclic macromolecules) (5). An important consideration for preparation of polymers with well-defined structures and low degrees of compositional heterogeneity is the choice of a suitable initiator.

In general, an appropriate initiator is a species which has approximately the same structure and reactivity as the propagating anionic species, ie, the pK_a of the conjugate acid of the propagating anion should correspond closely to the pK_a of the conjugate acid of the initiating species. If the initiator is too reactive, side reactions between the initiator and monomer can occur; if the initiator is not reactive enough, then the initiation reaction may be slow or inefficient.

The general relationship between monomer structural type, pK_a and appropriate initiating species is shown in Table 1. Those monomers which form the least stable anions, ie, which have the largest values of pK_a for the corresponding conjugate acids, are the least reactive monomers in anionic polymerization; in turn, these less reactive monomers require the use of the most reactive initiators as shown in Table 1. Thus, although the anionic polymerization of many heterocyclic monomers can be initiated with relatively weak bases such as alkoxides, hydroxides, and tertiary amines, in general the polymerization of vinyl monomers requires the use of alkali metals, aromatic radical anions, or organoalkali compounds.

Table 1. Relationships Between Monomer Reactivity, Carbanion Stability, and Suitable Initiators

Monomer type	pK_a^a		Initiators ^d
	In DMSO ^b	In H ₂ O ^c	
ethylene	56		RLi
dienes and	44		NH ₂ , RLi, RMt
styrenes	43		aromatic radical anions, ^e cumyl K, Mt,
acrylonitrile	32		RMgX
alkyl methacrylates, alkyl acrylates	30–31 (7)	27–28 (7)	fluorenyl, RAr ₂ C, ketyl radical anions ^f
vinyl ketones	26	19 (8)	
oxiranes	29–32	16–18 (9)	RO
thiiranes	17	12–13 (10)	
nitroalkenes	17	10–14 (11)	
siloxanes		10–14 (12,13)	RO, OH
β-lactones	12	4–5 (14)	RCOO
alkyl cyanoacrylates	12.8 (15)		HCO ₃ , H ₂ O
vinylidene cyanide	11	11 (15)	



^e For example, naphthalene radical anion with counterion (Li⁺, Na⁺, K⁺).

^a pK_a of the conjugate acid of the anionic propagating intermediate.

^b pK_a values in DMSO are from Ref. 6 unless noted in parentheses after the number.

^c Numbers in parentheses are references.

^d Mt refers generally to alkali metals (Li, Na, K, Rb, Cs).

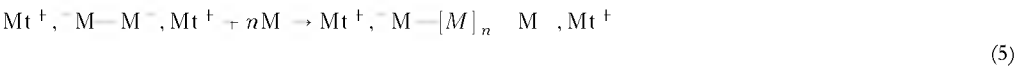
^f Ar₂CO[•]

Alkali Metals

The use of alkali metals for anionic polymerization of diene monomers is primarily of historical interest. A patent disclosure issued in 1911 (16) detailed the use of metallic sodium to polymerize isoprene and other dienes. Independently and simultaneously, the use of sodium metal to polymerize butadiene, isoprene, and 2,3-dimethyl-1,3-butadiene was described (17). Interest in alkali metal-initiated polymerization of 1,3-dienes culminated in the discovery (18) at Firestone Tire and Rubber Co. that polymerization of neat isoprene with lithium dispersion produced high *cis*-1,4-polyisoprene, similar in structure and properties to Hevea natural rubber (see ELASTOMERS, SYNTHETIC POLYISOPRENE; RUBBER, NATURAL).

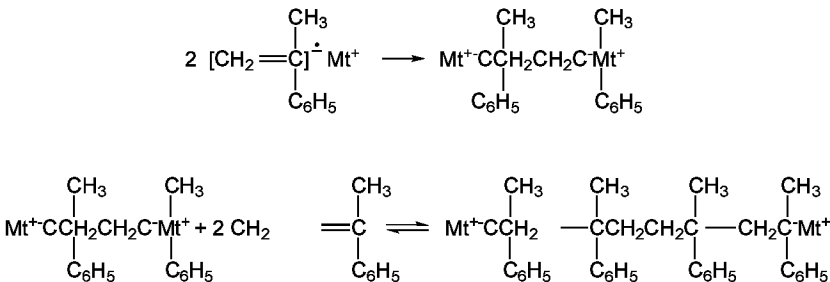
The mechanism of the anionic polymerization of styrenes and 1,3-dienes initiated by alkali metals has been described in detail (3,20) as shown in equations 3–5 where Mt represents an alkali metal and M is a monomer molecule. Initiation is a heterogeneous process occurring on the metal surface. The

first step is an electron-transfer reaction from the metal to the lowest unoccupied molecular orbital of an adsorbed monomer molecule to form radical anion intermediates which rapidly dimerize to form dianions (21–23). Monomer addition to these dianions forms adsorbed oligomers which eventually desorb and continue

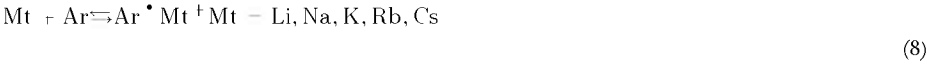


growth in solution. This heterogeneous initiation reaction continues to generate new active chain ends during the course of the subsequent propagation reactions. Consequently, there is little control of molecular weight and relatively broad molecular weight distributions ($M_w/M_n = 3-10$) have been reported for the soluble polymer obtained in bulk polymerizations (24). A polybutadiene polymer produced using sodium metal as initiator exhibits a high degree of branching and gel content (45%) combined with inhomogeneity in composition and molecular weight distribution (24,25).

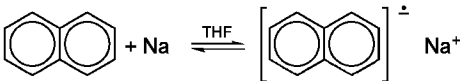
These reactions are useful for the preparation of homogeneous difunctional initiators from α -methylstyrene in polar solvents such as tetrahydrofuran. Because of the low ceiling temperature of α -methylstyrene ($T_c = 61^\circ\text{C}$) (26), dimers or tetramers can be formed depending on the alkali metal system, temperature, and concentration. Thus the reduction of α -methylstyrene by sodium potassium alloy produces the dimeric dianionic initiators in THF (27), while the reduction with sodium metal forms the tetrameric dianions as the main products (28). The structures of the dimer and tetramer correspond to initial tail-to-tail addition to form the most stable dianion as shown in equations 6 and 7 (28).



Aromatic Radical Anions. Many aromatic hydrocarbons react with alkali metals in polar aprotic solvents to form stable solutions of the corresponding radical anions as shown in equation 8 (3,20). These solutions can be analyzed by uv-visible spectroscopy and stored for further use. The unpaired electron is added to the lowest unoccupied molecular orbital of the aromatic hydrocarbon and a

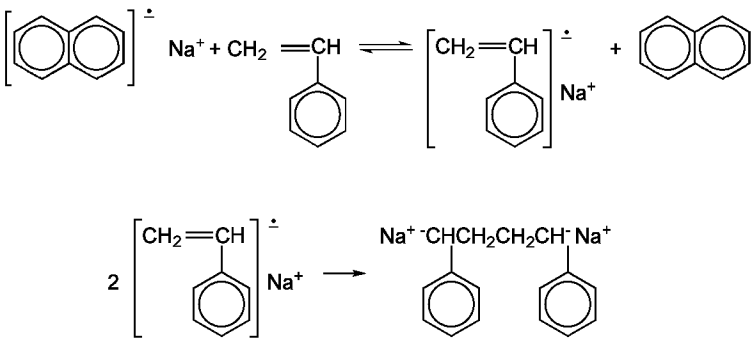


delocalized radical anion is formed as illustrated for sodium naphthalene in equation 9 (29–31). This oxidation–reduction reaction of the aromatic hydrocarbon with the metal is reversible; thus sodium metal and naphthalene are

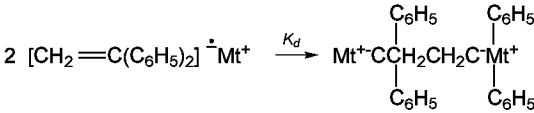


reformed when THF is removed. These radical anions can only be formed efficiently in polar aprotic solvents such as tetrahydrofuran (THF) and glycol methyl ethers (glymes). For example, although sodium naphthalene formation is 95% complete in tetrahydrofuran, this radical anion is formed in less than 1% yield in diethyl ether (20). For biphenyl, which has a lower electron affinity compared to naphthalene, only 20% of the corresponding radical anions are formed by sodium reduction in THF (20).

Sodium naphthalene [25398-08-7] and other aromatic radical anions react with monomers such as styrene by reversible electron transfer to form the corresponding monomer radical anions. Although the equilibrium (eq. 10)



between the monomer radical anion and the aromatic radical anion lies far to the left because of the low electron affinity of the monomer (3), this is an efficient initiation process because the resulting monomer radical anions dimerize rapidly (eqs. 11 and 12) with rate constants which approach the limits of diffusion control (20).



For example, the rate constants for dimerization of the radical anions of 1,1-diphenylethylene are 1.2×10^8 , 3.5×10^8 , 1×10^9 , and 3×10^9 L (mol · s) for the Li^+ , Na^+ , K^+ , and Cs^+ salts, respectively (20). Based on these kinetic results, it has been concluded that the addition of monomer to the monomer radical anion is of little importance in the electron-transfer initiation process (20).

Monomers which can be polymerized with aromatic radical anions include styrenes, dienes, epoxides, and cyclosiloxanes. Aromatic radical anions

which are too stable do not efficiently initiate polymerization of less reactive monomers; thus the anthracene radical anion cannot initiate styrene polymerization (20). In general, these aromatic radical anion initiators are formed and used in dipolar aprotic solvents; however, the addition of lithium naphthalene in THF to a benzene solution produces a finely divided suspension of lithium which reacts with styrene monomer to generate difunctional, narrow molecular weight distribution polymers (32). One of the consequences of the necessity to use polar solvents for aromatic radical anion initiators is that polydiene microstructure is high in 1,2- and 3,4-addition structures, ie, the high 1,4-stereospecificity observed with lithium in hydrocarbon solvent is lost in polar solvents such as tetrahydrofuran (1,33,34).

Alkyl lithium Compounds

Anionic polymerization of vinyl monomers can be effected with a variety of organometallic compounds; alkyl lithium compounds are the most useful class (1,33–35). A variety of simple alkyl lithium compounds are available commercially. Most simple alkyl lithium compounds are soluble in hydrocarbon solvents such as hexane and cyclohexane and they can be prepared by reaction of the corresponding alkyl chlorides with lithium metal. Methyl lithium [917-54-4] and phenyl lithium [591-51-5] are available in diethyl ether and cyclohexane–ether solutions, respectively, because they are not soluble in hydrocarbon solvents; vinyl lithium [917-57-7] and allyl lithium [3052-45-7] are also insoluble in hydrocarbon solutions and can only be prepared in ether solutions (38,39). Hydrocarbon-soluble alkyl lithium initiators are used directly to initiate polymerization of styrene and diene monomers quantitatively; one unique aspect of lithium-based initiators in hydrocarbon solution is that elastomeric polydienes with high 1,4-microstructure are obtained (1,24,33–37). Certain alkyl lithium compounds can be purified by recrystallization (ethyl lithium), sublimation (ethyl lithium, *t*-butyl lithium [594-19-4], isopropyl lithium [2417-93-8] or distillation (*sec*-butyl lithium) (40,41). Unfortunately, *n*-butyl lithium is noncrystalline and too high boiling to be purified by distillation (38). Since methyl lithium and phenyl lithium are crystalline solids which are insoluble in hydrocarbon solution, they can be precipitated into these solutions and then redissolved in appropriate polar solvents (42,43). Organometallic compounds of other alkali metals are insoluble in hydrocarbon solution and possess negligible vapor pressures as expected for salt-like compounds.

Simple alkyl lithium compounds are aggregated in solution, in the solid state, and even in the gas phase (38,44,45). The important differences between the various alkyl lithium compounds are their degrees of aggregation in solution (Table 2) and their relative reactivity as initiators for anionic polymerization of styrene and diene monomers. Alkyl lithium compounds are generally associated into dimers, tetramers, or hexamers in hydrocarbon solution (48). The degree of association is related to the steric requirements of the alkyl group, ie, the degree of association decreases as the steric requirements of the alkyl group increase. Models for the structures of the aggregated species have been deduced from analogous crystal structures determined in the solid state and are shown in Figure 1 (49,50). The tetramer can be described as interpenetrating tetrahedra of alkyl groups and lithium atoms or in terms of alkyl groups located above each of the triangular Li₃ faces of the tetrahedron formed by Li₄. In the hexameric aggregate, the lithium atoms occupy the apices of a distorted octahedron with six alkyl groups positioned above six of the eight octahedral Li₃ faces. Empty orbitals on lithium in the aggregates are available for interaction with Lewis bases.

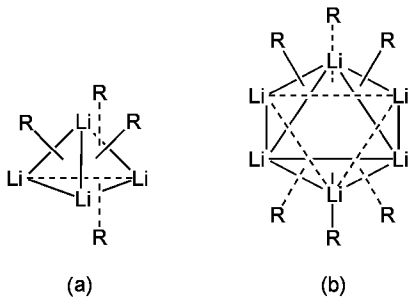


Fig. 1. Structures of alkyl lithium tetramers and hexamers: (a) tetrahedron of lithium alkyl groups; (b) arrangement of alkyl groups around octahedron of lithium atoms in hexamer.

Table 2. Association Numbers and Fractional Kinetic Orders for Alkyl lithium Initiators

RLi	Solvent	<i>N</i> ^a	Monomer	Reaction order ^b
<i>n</i> -C ₄ H ₉ Li	benzene	6	styrene	0.16
			butadiene	0.5–1.0
	cyclohexane	6.2	styrene	0.5–1.0
			butadiene	0.5–1.0
<i>sec</i> -C ₄ H ₉ Li	benzene	4	styrene	0.25
			butadiene	0.9
			isoprene	0.25
	cyclohexane	4	styrene	1.0
			butadiene	1.0
			isoprene	0.66–1.09
<i>t</i> -C ₄ H ₉ Li	cyclohexane		butadiene	1.0
			isoprene	0.2–0.7

^a Degree of association of RLi; Refs. 38, 40, and 41.
^b Kinetic order in [RLi] for initiation; Refs. 33, 36, 41, 46 and 47.

The kinetics of initiation reactions of alkyl lithium compounds often exhibit fractional kinetic order dependence on the total concentration of initiator as shown in Table 2. For example, the kinetics of the initiation reaction of *n*-butyl lithium with styrene monomer in benzene exhibit a first-order dependence on styrene concentration and a one-sixth order dependence on *n*-butyl lithium concentration as shown in equation 13, where *k_i* is the rate constant for

$$R_i = k_i K_d^{1/6} [C_4H_9Li]^{1/6} [M]$$

(13)

the initiation step and *K_d* is the dissociation equilibrium constant for the hexamer-unassociated RLi equilibrium (eq. 14). Since *n*-butyl lithium is aggre-



gated predominantly into hexamers in hydrocarbon solution, the fractional kinetic order dependency of the initiation process on total concentration of initiator is explained by assuming that the species which reacts with styrene monomer must be the unassociated form of the initiator and that this unassociated species may be formed by equilibrium dissociation processes. The kinetic order for *sec*-butyllithium-initiated polymerization of styrene is close to 0.25 in benzene solution. This result is consistent with the reaction of the unassociated form of the alkyllithium, since *sec*-butyllithium is associated predominantly into tetramers in benzene solution (see Table 2).

The use of aliphatic solvents causes profound changes in the observed kinetic behavior for the alkyllithium initiation reactions with styrene, butadiene, and isoprene, ie, the inverse correspondence between the reaction order dependence for alkyllithium and degree of organolithium aggregation is generally not observed (33,41,46). Also, initial rates of initiation in aliphatic solvents are several orders of magnitude less than those observed, under equivalent conditions, in aromatic solvents (41). Furthermore, pronounced induction periods are observed in aliphatic hydrocarbon solvents (41,46). It has been proposed that the initiation process in aliphatic hydrocarbons involves the direct reaction of monomer with aggregated organolithiums which then forms cross-associated species as shown in equation 15 (33,41,46). The induction periods are ascribed to the enhanced reactivity of the mixed (ie, cross-associated) aggregated species (46).



The relative reactivities of alkyllithiums as polymerization initiators are intimately linked to their degree of association. In the following the average degree of association in hydrocarbon solution, where known, is indicated in brackets after the alkyllithium (51). For styrene polymerization, the relative reactivity of alkyllithium initiators is menthyllithium [2] > *sec*-C₄H₉Li [4] > *i*-C₃H₇Li [4-6] > *i*-C₄H₉Li > *n*-C₄H₉Li [6] > *t*-C₄H₉Li [4]. For diene polymerization, menthyllithium [2] > *sec*-C₄H₉Li [4] > *i*-C₃H₇Li [4-6] > *t*-C₄H₉Li [4] > *i*-C₄H₉Li > *n*-C₄H₉Li [6]. In general, the less associated alkyllithiums are more reactive as initiators than the more highly associated species. Aromatic solvents tend to decrease the average degree of association and promote dissociation of aggregates which leads to initiation rates which are two to three orders of magnitude faster than in aliphatic solvents. Addition of Lewis bases such as ethers and amines tends to decrease the degree of aggregation of alkyllithium compounds (see Table 2) and accelerate rates of initiation (34).

Alkyllithium compounds are primarily used as initiators for polymerizations of styrenes and dienes (52). These initiators are too reactive for alkyl methacrylates and vinylpyridines. *n*-Butyllithium [109-72-8] is used commercially to initiate anionic homopolymerization and copolymerization of butadiene, isoprene, and styrene with linear and branched structures. Because of the high degree of association (hexameric), *n*-butyllithium-initiated polymerizations are often effected at elevated temperatures (>50°C) to increase the rate of initiation relative to propagation and thus to obtain polymers with narrower molecular weight distributions (53). Hydrocarbon solutions of this initiator are quite stable at room temperature for extended periods of time; the rate of decomposition per month is 0.06% at 20°C (39).

sec-Butyllithium [598-30-1] is the second most important organolithium initiator. It is used commercially to prepare styrene–diene block copolymers because it can initiate styrene polymerization rapidly compared to propagation so that even polystyrene blocks with relatively low molecular weights (10,000–15,000 g/mol) can be prepared with stoichiometric control and narrow molecular weight distributions (54). Hydrocarbon solutions of *sec*-butyllithium are thermally less stable than *n*-butyllithium solutions; the rate of decomposition is 1.4% per month at 20°C (39).

Quantitative Analysis of Alkyllithium Initiator Solutions. Solutions of alkyllithium compounds frequently show turbidity associated with the formation of lithium alkoxides by oxidation reactions or lithium hydroxide by reaction with moisture. Although these species contribute to the total basicity of the solution as determined by simple acid titration, they do not react with allylic and benzylic chlorides or ethylene dibromide rapidly in ether solvents. This difference is the basis for the double titration method of determining the amount of active carbon-bound lithium reagent in a given sample (55,56). Thus the amount of carbon-bound lithium is calculated from the difference between the total amount of base determined by acid titration and the amount of base remaining after the solution reacts with either benzyl chloride, allyl chloride, or ethylene dibromide.

Copolymerization Initiators. The copolymerization of styrene and dienes in hydrocarbon solution with alkyllithium initiators produces a tapered block copolymer structure because of the large differences in monomer reactivity ratios for styrene (*r_s* < 0.1) and dienes (*r_d* > 10) (1,33,34). In order to obtain random copolymers of styrene and dienes, it is necessary to either add small amounts of a Lewis base such as tetrahydrofuran or an alkali metal alkoxide (MtOR, where Mt = Na, K, Rb, or Cs). In contrast to Lewis bases which promote formation of undesirable vinyl microstructure in diene polymerizations (57), the addition of small amounts of an alkali metal alkoxide such as potassium amyloxide ([ROK] [Li] = 0.08) is sufficient to promote random copolymerization of styrene and diene without producing significant increases in the amount of vinyl microstructure (58,59).

Difunctional Initiators. These initiators are of considerable interest for the preparation of triblock copolymers, telechelic polymers, and macrocyclic polymers. Although triblock copolymers can be prepared with monofunctional initiators using a three-step, sequential monomer addition process, with difunctional initiators they can be formed in a more efficient two-step process (34,60,61). Difunctional initiators also provide a methodology to prepare new triblock copolymers which cannot be prepared by the three-step, sequential monomer addition route because the chain ends formed from the first monomer are too stable to initiate the polymerization of the second monomer; for example, a difunctional initiator can be used for the direct synthesis of poly(ethylene oxide)-*block*-polystyrene-*block*-poly(ethylene oxide) (61). Difunctional initiators provide direct, efficient methods for the formation of α, ω-difunctional polymers, ie, telechelic polymers (62), by termination reactions of the polymeric α, ω-dianions with electrophilic functionalization agents. Analogously, termination of the α, ω-dianions with a difunctional, electrophilic coupling agent under high dilution conditions promotes intramolecular cyclization reactions to form macrocyclic polymers (63).

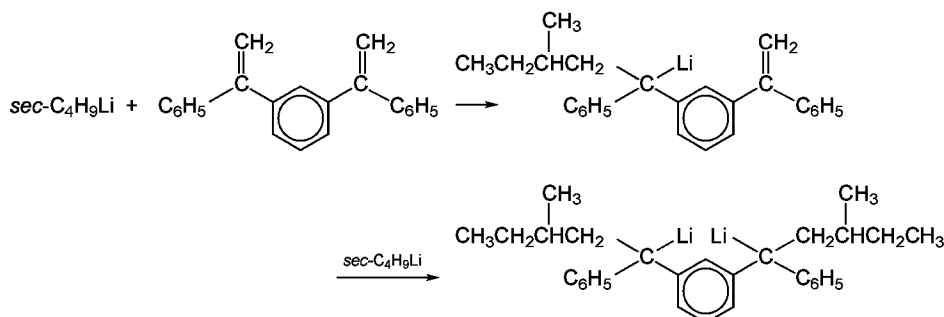
Aromatic radical anions, such as lithium naphthalene or sodium naphthalene, are efficient difunctional initiators (eqs. 6,7) (3,20,64). However, the necessity of using polar solvents for their formation and use limits their utility for diene polymerization, since the unique ability of lithium to provide high 1,4-polydiene microstructure is lost in polar media (1,33,34,57,63,64). Consequently, a significant research challenge has been to discover a hydrocarbon-soluble dilithium initiator which would initiate the polymerization of styrene and diene monomers to form monomodal α, ω-dianionic polymers at rates which are faster or comparable to the rates of polymerization, ie, to form narrow molecular weight distribution polymers (61,65,66).

The methodology for preparation of hydrocarbon-soluble, dilithium initiators is generally based on the reaction of an aromatic divinyl precursor with two moles of butyllithium. Unfortunately, because of the tendency of organolithium chain ends in hydrocarbon solution to associate and form electron-deficient dimeric, tetrameric, or hexameric aggregates (see Table 2) (33,38,44,67), attempts to prepare dilithium initiators in hydrocarbon media have generally resulted in the formation of insoluble, three-dimensionally associated species (34,66,68–72). These precipitates are not effective initiators because of their heterogeneous initiation reactions with monomers which tend to result in broader molecular weight distributions (*M_w* / *M_n* > 1.1) (68,70,72). Soluble analogues of these difunctional initiators have been prepared either by addition of small amounts of weakly basic additives such as triethylamine (73) or anisole (74) which have relatively minor effects on diene microstructure (37). Another method to solubilize these initiators is to use a seeding technique, whereby small amounts of diene monomer are added to form a hydrocarbon-soluble, oligomeric dilithium-initiating species (69,75).

The stoichiometric reaction of *m*-diisopropenylbenzene [3748-13-8] with two moles of *sec*-butyllithium in the presence of triethylamine has been reported to produce a useful, hydrocarbon-soluble dilithium initiator because of the low ceiling temperature of the monomer (78,79) which is analogous in structure to α-methylstyrene; however, other studies suggest that oligomerization occurs to form initiators with functionalities higher than two (80).

Analogously, the use of *m*-divinylbenzene [108-57-6] has been reported (76). However, oligomerization occurs on treatment of divinylbenzene with butyllithium resulting in initiators with functionalities greater than two (68). From a commercial perspective this oligomerization and lack of precise functionality control is not necessarily a problem and useful multifunctional initiators have been prepared from the reaction of butyllithium with varying amounts of divinylbenzene (commercial divinylbenzene contains 22% *meta*, 11% *para*, and 66% *o*-, *m*-, and *p*-ethylvinylbenzene) (77) often in the presence of styrene or diene monomer to provide solubility (seeding technique).

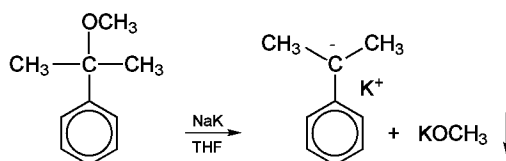
Although a plethora of divinyl aromatic compounds have been investigated as precursors for hydrocarbon-soluble dilithium initiators (68), the only system which has been demonstrated to produce a hydrocarbon-soluble dilithium initiator is based on 1,3-bis(1-phenylethenyl)benzene (60,81–85). The addition reaction of *sec*-butyllithium with 1,3-bis(1-phenylethenyl)benzene [34241-86-6] (eq. 16) proceeds rapidly and efficiently to produce the corresponding dilithium species in toluene (86) or in cyclohexane (82). This dilithium initiator is not only soluble in hydrocarbon media such as cyclohexane, benzene, and toluene (even at -20°C) (84), but also functions as an efficient difunctional initiator for the preparation of homopolymers and triblock copolymers with relatively narrow molecular weight distributions (81–83). However, it is necessary to add a small amount of Lewis base or two equivalents of lithium *sec*-butoxide to produce narrow, monomodal molecular weight distributions. Lithium *sec*-butoxide is the preferred additive, since high 1,4-polybutadienes are obtained (60).



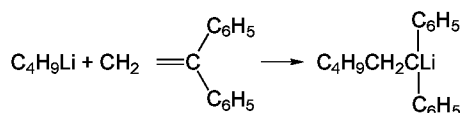
Functionalized Initiators. The use of alkyllithium initiators which contain functional groups provides a versatile method for the preparation of end functionalized polymers and macromonomers. For a living anionic polymerization, each functionalized initiator molecule produces one macromolecule with the functional group from the initiator residue at one chain end and the active carbanionic propagating species at the other chain end. Thus, in contrast to most functionalization procedures which involve post-polymerization termination reactions with electrophilic reagents (87), the use of a functionalized initiator retains the anionic chain end and the ability to prepare block and star-branched polymers with the functional group at the initiating end. For example, dimethylaminopropyllithium can be prepared in hydrocarbon solution and has been used to prepare polystyrenes and polydienes with tertiary amine end group functionality (88,89). However, many functional groups such as hydroxyl, carboxyl, phenol, and primary amine are not stable in the presence of reactive dienyllithium and styryllithium chain ends. Therefore, it is necessary to convert these functional groups into suitable derivatives, ie, protected groups, which must be stable to the carbanionic chain ends and which can be removed readily after polymerization is completed (90). Examples of these types of protected functional initiators include the hydroxyl-protected initiator, 6-lithiohexyl acetaldehyde acetal (91) and a primary amine-protected initiator, 4-bis(trimethylsilyl)aminophenyllithium (92).

Other Initiators

Cumyl Potassium. Cumyl potassium [3003-91-6] ($\text{p}K_a > 43$ based on toluene) (6) is another useful initiator for anionic polymerization of a variety of monomers, including styrenes, dienes, methacrylates, and epoxides. This carbanion is readily prepared from cumyl methyl ether as shown in equation 17 (93). It is necessary to remove the potassium methoxide salt which precipitates from the solution; cooling to low temperature prior to filtration is recommended. The concentration of active initiating species can be determined by titration with standardized acid or by using this initiator with a known amount of styrene monomer and measuring the number average molecular weight of the polymer assuming that one initiator moiety produces one polystyrene macromolecule. This initiator is generally used at low temperatures in a polar solvent such as THF, which limits the microstructure of polydienes to low 1,4-contents.

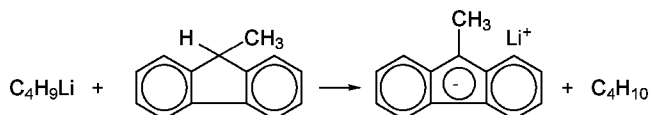


1,1-Diphenylmethylcarbanions. The carbanions based on diphenylmethane ($\text{p}K_a = 32$) (6) are useful initiators for vinyl and heterocyclic monomers, especially alkyl methacrylates at low temperatures (94,95). Addition of lithium chloride or lithium *tert*-butoxide has been shown to narrow the molecular weight distribution and improve the stability of active centers for anionic polymerization of both alkyl methacrylates and *tert*-butyl acrylate (96,97). Surprisingly, these more stable carbanions can also efficiently initiate the polymerization of styrene and diene monomers (98). Diphenylmethylolithium [881-42-5] can be prepared by the metalation reaction of butyllithium with diphenylmethane; in addition, the adduct of butyllithium and 1,1-diphenylethylene is conveniently prepared in either hydrocarbon or polar solvents such as THF as shown in equation 18.



This reaction can also be utilized to prepare functionalized initiators by reaction of butyllithium with a substituted 1,1-diphenylethylene derivative. For example, polymers end functionalized with primary amine, tertiary amine, phenol, and bis(phenol) groups have been prepared in essentially quantitative yield by using the reaction of butyllithium with the corresponding substituted (or protected) 1,1-diphenylethylene (87).

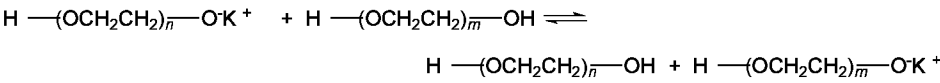
Fluorenyl Carbanions. Salts of fluorene ($\text{p}K_a = 22.6$) (6) are more hindered and less reactive than many other organometallic initiators. These carbanions can be readily formed by reaction with alkali metal derivatives as shown in equation 19 for 9-methylfluorene (99). Carbanion salts of 9-methylfluorene are preferable to fluorene, since the latter generate chain ends which retain reactive, acidic fluorenyl hydrogens which can participate in chain-transfer reactions (100,101). Fluorenyl salts are useful initiators for the polymerization of alkyl methacrylates, epoxide, and thiirane monomers.



Enolate Initiators. In principle, ester enolate anions should represent the ideal initiators for anionic polymerization of alkyl methacrylates. Although general procedures have been developed for the preparation of a variety of alkali metal enolate salts, many of these compounds are unstable except at low temperatures (67,102,103). Useful initiating systems for acrylate polymerization have been prepared from complexes of ester enolates with alkali metal alkoxides (104,105).

Alkoxide-Type Initiators. Using the guide that an appropriate initiator should have approximately the same structure and reactivity as the propagating anionic species (see Table 1), alkoxide, thioalkoxide, carboxylate, and silanolate salts would be expected to be useful initiators for the anionic polymerization of epoxides, thiiranes, lactones, and siloxanes, respectively (106–108). Thus low molecular weight poly(ethylene oxide) can be prepared

using catalytic amounts of potassium hydroxide in the presence of water, alcohol, or glycol (109) (see POLYETHERS, ETHYLENE OXIDE POLYMERS). The presence of hydroxylic groups is not a problem because of the fast proton-transfer equilibria between active chain ends as shown in equation 20 (109). In order to prepare higher molecular



weight poly(ethylene oxide) (10^5-10^6 g mol), catalysts such as strontium carbonate ($0.1-0.4^\circ$ o H_2O), calcium amide, or calcium amide alkoxide are required (109). Thiiranes are more reactive than epoxides and their polymerization can even be initiated with amines and phosphines (110). For β -propiolactone, the propagating anion is a carboxylate, while for ϵ -caprolactone it is an alkoxide (106,110). Thus, although carboxylates can be used to initiate polymerization of β -propiolactone, alkoxides are necessary for polymerization of less reactive ϵ -caprolactone. Hydroxides, alkoxides, and silanolates (R_3SiO^-) are effective initiators for the polymerization of cyclosiloxanes (108). In general, the larger counterions give rise to more active catalysts ($\text{Li}^+ < \text{Na}^+ < \text{K}^+ < \text{Rb}^+ < \text{Cs}^+ \approx \text{R}_4\text{N}^+$) (108).

Health and Safety Factors

Hydrocarbon solutions of alkyllithium compounds are air and moisture sensitive and should be either handled in an inert atmosphere or by using syringes using recommended procedures for handling air-sensitive compounds (111). Alkyllithium reagents react with acidic compounds that contain reactive hydrogens such as water, alcohols, phenols, acids, and even primary and secondary amines (38). The reaction of butyllithium with water produces butane and lithium hydroxide, which can lead to spontaneous ignition in the presence of oxygen (39). Contact of alkyllithium solutions with air does not generally lead to spontaneous ignition; however, if large surface areas are formed, for example in a spill, spontaneous ignition can occur. Carbon dioxide fire extinguishers must not be used because carbon dioxide reacts exothermically with alkyllithium compounds. It is prudent to have an all-purpose fire extinguisher available when working with these organometallic compounds. Suitable fire-extinguishing chemicals include powdered limestone and powders containing sodium chloride and sodium bicarbonate (39).

BIBLIOGRAPHY

1.

M. Morton, *Anionic Polymerization: Principles and Practice*, Academic Press, Inc., New York, 1982.

2.

M. Szwarc and M. Van Beylen, *Ionic Polymerization and Living Polymers*, Chapman and Hall, New York, 1993.

3.

M. Szwarc, *Carbanions, Living Polymers and Electron Transfer Processes*, Wiley-Interscience, New York, 1968.

4.

R. P. Quirk and B. Lee, *Polym. Int.* **27**, 359 (1992).

5.

P. Rempp, E. Franta, and J.-E. Herz, *Adv. Polym. Sci.* **86**, 145 (1988).

6.

F. G. Bordwell, *Account. Chem. Res.* **21**, 456 (1988).

7.

F. G. Bordwell and D. J. Algrim, *J. Am. Chem. Soc.* **110**, 2964 (1988).

8.

Y. Chiang, A. J. Kresge, and Y. S. Tang, *J. Am. Chem. Soc.* **106**, 460 (1984).

9.

W. Reeve, C. M. Erikson, and P. F. Aluotto, *Can. J. Chem.* **57**, 2747 (1979).

10.

A. Ohno and S. Oae, in S. Oae, ed., *Organic Chemistry of Sulfur*, Plenum Press, New York, 1977, p. 128.

11.

R. G. Pearson and R. L. Dillon, *J. Am. Chem. Soc.* **75**, 2439 (1953).

12.

L. Allred, E. G. Rochow, and F. G. A. Stone, *J. Inorg. Nucl. Chem.* **2**, 416 (1956).

13.

H. Arm, K. Hochstrasser, and P. W. Schindler, *Chimia* **28**, 237 (1974).

14.

G. Kortum, W. Vogel, and K. Andrussov, *Dissociation Constants of Organic Acids in Aqueous Solution*, Butterworths, London, 1961.

15.

F. G. Bordwell and H. E. Fried, *J. Org. Chem.* **46**, 4327 (1981).

16.

Brit. Pat. 24,790 (1910) F. E. Mathews and E. H. Strange.

17.

C. H. Harries, *Leibigs Ann. Chem.* **383**, 184, 190 (1911); **U. S. Pat. 1,058,056** (Apr. 8, 1913).

18.

F. W. Stavely and co-workers, *Ind. Eng. Chem.*, 778 (1956).

19.

F. C. Foster and J. L. Binder, in *Handling and Uses of Alkali Metals*, Advances in Chemistry series no. 19, American Chemical Society, Washington, D.C., 1957, p. 26.

20.

M. Szwarc, *Adv. Polym. Sci.* **49**, 1 (1983).

21.

H. C. Wang, G. Levin, and M. Szwarc, *J. Am. Chem. Soc.* **100**, 3969 (1978).

22.

H. C. Wang, G. Levin, and M. Szwarc, *J. Phys. Chem.* **83**, 785 (1979).

23.

M. Szwarc, in F. E. Bailey, Jr., ed., *Initiation of Polymerization*, ACS Symposium series 212, American Chemical Society, Washington, D.C., 1983, p. 419.

24.

E. W. Duck and J. M. Locke, in W. M. Saltman, ed., *The Stereo Rubbers*, Wiley-Interscience, New York, 1977, p. 139.

25.

H. Blumel, *Rubber Chem. Technol.* **37**, 408 (1964).

26.

K. J. Ivin and W. K. Busfield, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, Vol. 12, 2nd ed., John Wiley & Sons, Inc., New York, 1988, p. 555.

27.

C. L. Lee, J. Smid, and M. Szwarc, *J. Phys. Chem.* **66**, 904 (1962).

28.

D. H. Richards and R. L. Williams, *J. Polym. Sci.* **11**, 89 (1973).

29.

A. Streitwieser, Jr., *Molecular Orbital Theory for Organic Chemists*, John Wiley & Sons, Inc., New York, 1961, p. 159.

30.

N. L. Holy, *Chem. Rev.* **74**, 243 (1974).

31.

F. Gerson and W. Huber, *Account. Chem. Res.* **20**, 85 (1987).

32.

R. P. Quirk and P. Cheng, *Macromolecules* **19**, 1291 (1986).

33.

R. N. Young, R. P. Quirk, and L. J. Fetters, *Adv. Polym. Sci.* **56**, 1 (1984).

34.

S. Bywater, in Ref. 26, Vol. 2, p. 1.

35.

R. P. Quirk and V. Monroy, *Ind. J. Technol.* **31**, 204 (1993).

36.

M. Morton and L. J. Fetters, *Rubber Chem. Technol.* **48**, 359 (1975).

37.

S. Bywater, in G. C. Eastmond and co-workers, eds., *Comprehensive Polymer Science*, Vol. 3, *Chain Polymerization I*, Pergamon Press, Elmsford, N.Y., 1989, p. 433.

38.

B. J. Wakefield, *The Chemistry of Organolithium Compounds*, Pergamon Press, Elmsford, N.Y., 1974.

39.

R. J. Bauer, in *Ullmann's Encyclopedia of Industrial Chemistry*, Vol. A15, 5th ed., VCH Verlagsgesellschaft, Weinheim, Germany, 1990, p. 393.

40.

H. L. Lewis and T. L. Brown, *J. Am. Chem. Soc.* **92**, 4664 (1970).

41.

S. Bywater and D. J. Worsfold, *J. Organometal. Chem.* **10**, 1 (1967).

42.

T. L. Brown and M. T. Rogers, *J. Am. Chem. Soc.* **79**, 1859 (1957).

43.

G. R. Brubaker and P. Beak, *J. Organometal. Chem.* **136**, 147 (1977).

44.

J. L. Wardell, in G. Wilkinson, R. G. A. Stone, and E. W. Abel, eds., *Comprehensive Organometallic Chemistry: The Synthesis, Reactions and Structures of Organometallic Compounds*, Vol. 1, Pergamon Press, Oxford, U.K., 1982, p. 43.

45. T. L. Brown, *Pure Appl. Chem.* **23**, 447 (1970).

46. J. E. L. Roovers and S. Bywater, *Macromolecules* **1**, 328 (1968).

47. S. Bywater in C. H. Bamford and C. F. H. Tipper, eds., *Comprehensive Chemical Kinetics*, Amsterdam, the Netherlands, 1976, p. 1.

48. T. L. Brown, *Account. Chem. Res.* **1**, 23 (1968).

49. E. Weiss, *Angew. Chem., Int. Ed.* **32**, 1501 (1993).

50. K. Gregory, P. von r. Schleyer, and R. Snaith, *Adv. Organometal. Chem.* **37**, 47 (1992).

51. H. L. Hsieh and W. H. Glaze, *Rubber Chem. Technol.* **43**, 22 (1970).

52. H. L. Hsieh, R. C. Farrar and K. Udiipi, *CHEMTECH*, 626 (1981).

53. H. L. Hsieh and O. F. McKinney, *Polym. Lett.* **4**, 843 (1966).

54. G. Holden and N. R. Legge, in N. R. Legge, G. Holden, and H. E. Schroeder, eds., *Thermoplastic Elastomers: A Comprehensive Review*, Hanser, Munich, 1987, p. 47.

55. H. Gilman and F. K. Cartledge, *J. Organometal. Chem.* **2**, 447 (1964).

56. T. R. Crompton, *Comprehensive Organometallic Analysis*, Plenum Press, New York, 1987, p. 181.

57. T. Antkowiak and co-workers, *J. Polym. Sci., Part A-1* **10**, 1319 (1972).

58. C. F. Wofford and H. L. Hsieh, *J. Polym. Sci., Part A-1* **7**, 461 (1969).

59. *Ibid.*, p. 449.

60. R. P. Quirk and J.-J. Ma, *Polym. Int.* **24**, 197 (1991).

61. G. Riess and G. Hurtrez, in Ref. 34, p. 324.

62. E. J. Goethals, ed., *Telechelic Polymers: Synthesis and Applications*, CRC Press, Boca Raton, Fla., 1989.

63. P. Rempp, C. Strazielle, and P. Lutz, in Ref. 26, Vol. 9, p. 183.

64. M. J. M. Abadie and L. Satibi, *Eur. Polym. J.* **23**, 423 (1987).

65. R. Jerome, R. Fayt, and T. Ouhadi, *Prog. Polym. Sci.* **10**, 87 (1984).

66. M. Fontanille, in Ref. 37, p. 365.

67. P. G. Williard, in B. M. Trost, ed., *Comprehensive Organic Synthesis*, Vol. 1, Part 1, Pergamon Press, Oxford, U.K., 1991.

68. F. Bandermann, H.-D. Speikamp, and L. Weigel, *Makromol. Chem.* **186**, 2017 (1985).

69. P. Guyot and co-workers, *Polymer* **22**, 1724 (1981).

70. P. Guyot and co-workers, *Polymer* **23**, 73 (1982).

71. I. Obriot, J. C. Favier, and P. Sigwalt, *Polymer* **28**, 2093 (1987).

72. A. Soum, M. Fontanille, and A. Aboudalle, *Polymer* **29**, 1528 (1988).

73. R. P. Foss, H. W. Jacobson, and W. H. Sharkey, *Macromolecules* **10**, 287 (1977).

74. L. J. Fetters and M. Morton, *Macromolecules* **2**, 453 (1969).

75. L. H. Tung, G. Y.-S. Lo, and D. E. Beyer, *Macromolecules* **11**, 616 (1978).

76. P. Lutz, E. Franta, and P. Rempp, *C. R. Acad. Sci., Ser. C* **283**, 123 (1976).

77. B. J. Bauer and L. J. Fetters, *Rubber Chem. Technol.* **51**, 406 (1978).

78. G. Beinert and co-workers, *Makromol. Chem.* **179**, 551 (1978).

79. P. Lutz, E. Franta, and P. Rempp, *Polymer* **23**, 1953 (1982).

80. G. G. Cameron and G. M. Buchan, *Polymer* **20**, 1129 (1979).

81. L. H. Tung and G. Y. Lo, in J. Lal and J. E. Mark, eds., *Advances in Elastomers and Rubber Elasticity*, Plenum Press, New York, 1986, p. 129.

82. A. D. Broske and co-workers, in T. E. Hogen-Esch and J. Smid, eds., *Recent Advances in Anionic Polymerization*, Elsevier, New York, 1987, p. 363.

83. T. E. Long and co-workers, *J. Polym. Sci., Polym. Chem.* **27**, 4001 (1989).

84. G. Schulz and H. Hucker, *Angew. Chem. Int. Ed.* **19**, 219 (1980).

85. T. Bastelberger and H. Hucker, *Angew. Makromol. Chem.* **125**, 53 (1984).

86. E. Leitz and H. Hucker, *Makromol. Chem.* **184**, 1893 (1983).

87. R. P. Quirk, in S. L. Aggarwal and S. Russo, eds., *Comprehensive Polymer Science*, 1st suppl., Pergamon Press, Oxford, U.K., 1992, p. 83.

88. N. S. Davidson and co-workers, *Macromolecules* **21**, 112 (1988).

89. H. J. Taunton and co-workers, *Macromolecules* **23**, 571 (1990).

90. S. Nakahama and A. Hirao, *Prog. Polym. Sci.* **15**, 299 (1990).

91. D. N. Schulz, A. F. Halasa, and A. E. Oberster, *J. Polym. Sci., Polym. Chem.* **12**, 153 (1974).

92. D. N. Schulz and A. F. Halasa, *J. Polym. Sci., Polym. Chem.* **15**, 2401 (1977).

93. K. Ziegler and H. Dislich, *Chem. Ber.* **90**, 1107 (1957).

94. G. D. Andrews and L. R. Melby, in B. M. Culbertson and C. U. Pittman, Jr., eds., *New Monomers and Polymers*, Plenum Press, New York, 1984, p. 357.

95. B. C. Anderson and co-workers, *Macromolecules* **14**, 1599 (1981).

96. R. Fayt and co-workers, *Macromolecules* **20**, 1442 (1987).

97. M. Janata, L. Lochmann, and A. H. E. Moller, *Makromol. Chem.* **194**, 625 (1993).

98. R. P. Quirk and L.-F. Zhu, *Brit. Polym. J.* **23**, 47 (1990).

99. R. P. Quirk and G. P. Bidinger, *Polym. Bull.* **22**, 63 (1989).

100. D. Lassalle, S. Boileau, and P. Sigwalt, *Eur. Polym. J.* **13**, 599 (1977).

101. A. Deffieux, P. Sigwalt, and S. Boileau, *Eur. Polym. J.* **20**, 77 (1984).

102. D. Seebach, *Angew. Chem. Int. Ed.* **27**, 1624 (1988).

103. G. Boche, K. Harms, and M. Marsch, *Makromol. Chem., Macromol. Symp.* **67**, 97 (1993).

104. P. Vlcek and L. Lochmann, *Makromol. Chem., Macromol. Symp.* **67**, 111 (1993).

105. D. Kinkel and co-workers, *Makromol. Chem. Macromol. Symp.* **60**, 315 (1992).

106. S. Slomkowski and A. Duda, in D. J. Brunelle, ed., *Ring-Opening Polymerization*, Hanser, New York, 1993, p. 87.

107. K. J. Ivin and T. Saegusa, eds., *Ring-Opening Polymerization*, Vols. 1 and 2, Elsevier, New York, 1984.

108. T. C. Kendrick, B. M. Parbhoo, and J. W. White, in Ref. 37, p. 459.

109. N. Clinton and P. Matlock, in Ref. 26, Vol. 6, p. 225.

110. T. Tsuruta and Y. Kawakami, in Ref. 37, p. 457.

111. D. F. Shriver and M. A. Drezdson, *The Manipulation of Air-Sensitive Compounds*, 2nd ed., Wiley-Interscience, New York, 1986.

Roderic P. Quirk
University of Akron

Victor M. Monroy
General Tire, Inc.

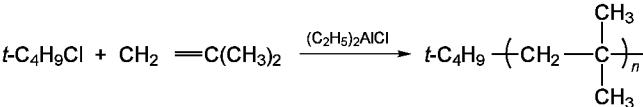
the left for the growing polymer chain ends may result in very low growing center concentration and therefore in negligible polymerization rates.

Solvent Polarity and Temperature. The dielectric constant and polarizability are of little predictive value for the selection of solvents relative to polymerization rates and behavior. In spite of the similarity of the dielectric constants of CH₂Cl₂, CH₃Cl, and C₂H₅Cl these solvents yield quite different isobutylene polymerization rates that decrease in the same order.

In cationic polymerization of vinyl monomers chain transfer is the most significant chain-breaking process. The activation energy of chain transfer is higher than that of propagation; consequently, the molecular weight of the polymer increases with decreasing temperature. Intramolecular alkylation, an undesirable side reaction in the polymerization of styrenic monomers and in the polymerization of isobutylene by the inifer technique, can also be eliminated by lowering the temperature and solvent polarity. However, opposite results were reported for SnBr₄ in the polymerization of α-methylstyrene, eg, intramolecular alkylation occurred using toluene but was absent using CH₂Cl₂ (6). Evidently every system has to be examined independently.

Controlled Initiation

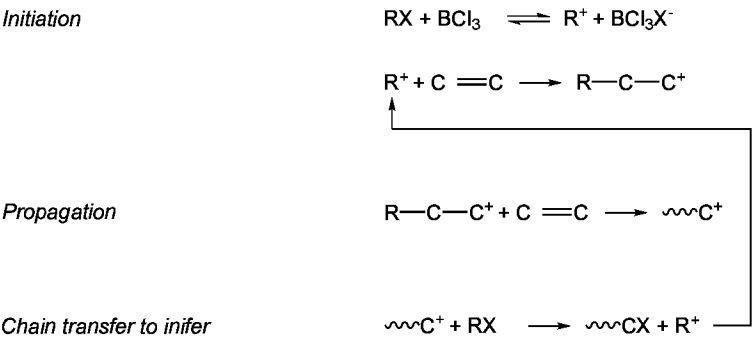
Initiation by a carbocation source provides control of the head group (controlled initiation) when used in conjunction with a Friedel-Crafts acid (for instance (C₂H₅)₃Al, (CH₃)₃Al, (C₂H₅)₂AlCl, BCl₃ for isobutylene, or I₂ and zinc halides for vinyl ethers) where chain transfer to monomer is absent or negligible, or in the presence of a proton trap to abort chain transfer to monomer (11). That is, initiation from tertiary, allylic, and benzylic halides gives rise to macromolecules carrying tertiary, allylic, and benzylic head groups. Initiation by halogens results in head groups carrying the halogen. Controlled initiation, however, is achieved only when polymer formation from adventitious protic impurities is also absent or negligible.



Polymer formation from protic impurities can be minimized by increasing the concentration of initiator or can be eliminated by the use of proton traps (12), eg, 2,6-di-*tert*-butylpyridine (DTBP) or similar hindered pyridines (for instance 2,6-di-*tert*-butyl-4-methylpyridine) which exhibit extraordinary specificity toward protons owing to their very high basicity coupled with nonnucleophilicity due to steric hindrance (13). Accordingly, they react with HCl but not with BF₃ or CH₃I. The specificity toward protons has been exploited to obtain high efficiency block and graft copolymer formation by aborting counteranion assisted chain transfer in the polymerization of α-methylstyrene and isobutylene (14,15), and to achieve living homo- and block copolymerization of isobutylene and styrenes (16).

When using a cation source in conjunction with a Friedel-Crafts acid the concentration of growing centers is most often difficult to measure and remains unknown. By the use of stable carbocation salts (for instance trityl and tropylium hexachloroantimonate) the uncertainty of the concentration of initiating cations is eliminated. Due to the highly reproducible rates, stable carbocation salts have been used in kinetic studies. Their use, however, is limited to cationically fairly reactive monomers (eg, *N*-vinylcarbazole, *p*-methoxystyrene, alkyl vinyl ethers) since they are too stable and therefore ineffective initiators of less reactive monomers, such as isobutylene, styrene, and dienes.

The Inifer Method. A special case of controlled initiation is the inifer method (17). The word inifer (from *i*nitiator *tr*ansfer agents) describes compounds that function simultaneously as initiators and as chain-transfer agents. Chain transfer to inifer regenerates R⁺. The inifer technique provided the first carbocationic route toward the synthesis of telechelic (α,ω functional) polyisobutylenes and more recently telechelic poly(*p*-chlorostyrenes) (18). To prepare telechelic products chain transfer to monomer must be absent, and with BCl₃ as coinitiator this requirement is fulfilled.



Direct Initiation

The mechanism of initiation in cationic polymerization using Friedel-Crafts acids appeared to be clarified by the discovery that most Friedel-Crafts acids, particularly halides of boron, titanium, and tin, require an additional cation source to initiate polymerization. Evidence has been accumulating, however, that in many systems Friedel-Crafts acids alone are able to initiate cationic polymerization. The polymerization of isobutylene for instance can be initiated, reportedly even in the absence of an added initiator, by AlBr₃ or AlCl₃ (19), TiCl₄ (20), AlC₂H₅Cl₂ (21), and BCl₃ (22). Three fundamentally different theories have been presented to explain the still controversial existence of direct initiation. Halometalation is proposed by Sigwalt and Olah (23). In the presence of excess Friedel-Crafts acid the formed metalloorganic compound may ionize or eliminate HCl, a conventional cationogen. Self-ionization of the Friedel-Crafts acid has been suggested to explain direct initiation (19,20). Allylic self-initiation may also explain results with olefins possessing an allylic hydrogen (24).

All three theories imply that the polymerization system is free of protogenic impurities. Although direct initiation by metal halides has been postulated with the above Friedel-Crafts acids, it was proven only for aluminum halides (19) and more recently for BCl₃ (22). With TiCl₄, attempts have been made to observe the corresponding intermediates by ¹H nmr but without success, which was explained by the known instability of the organotitanium compounds (20). Kinetic investigation of polymerizations by BCl₃ (22) suggests that initiation is by haloboration according to the Sigwalt-Olah theory. Initiation by I₂ in the polymerization of vinyl ethers can be visualized similarly, ie, a 1,2-diiodide is formed first that is subsequently activated by excess I₂ (25).

Ring-Opening Polymerization

Many initiating systems used in the cationic polymerization of vinyl monomers can also be used to initiate ring-opening polymerization of cyclic monomers such as cyclic ethers, acetals, lactams, lactones, and siloxanes. Polymerization of cyclic monomers may involve different types of ionic as well as covalent growing species. Under certain conditions termination processes may be absent. The polymerization of cyclic monomers, however, is almost always complicated by inter- and intramolecular chain transfer to polymer, resulting in cyclic oligomer formation. The extent of cyclic oligomer formation can be minimized in the polymerization of epoxides by the recently discovered activated monomer mechanism (26). The polymerization is carried out in the presence of alcohol at very low monomer concentration by continuous monomer feeding.

Cyclic ether and acetal polymerizations are also important commercially. Polymerization of tetrahydrofuran is used to produce polyether diol, and polyoxymethylene, an excellent engineering plastic, is obtained by the ring-opening polymerization of trioxane with a small amount of cyclic ether or acetal comonomer to prevent depolymerization (see ACETAL RESINS; POLYETHERS, TETRAHYDROFURAN).

Living Cationic Polymerization

A variety of initiating systems have been described that allow not only controlled initiation but also controlled propagation in the polymerization of vinyl monomers. In these living polymerization systems chain breaking (chain transfer and irreversible termination) is absent. The first example of living carbocationic polymerization, the polymerization of isobutyl vinyl ether [109-53-5] with a mixture of hydrogen iodide and iodine (HI I₂) was discovered in 1984 (27). Since then the scope has been rapidly expanded to different vinyl ethers, propenyl ethers and other cationically highly reactive monomers, such as *N*-vinylcarbazole [1484-13-5] and *p*-methoxystyrene [637-69-4], and to other initiating systems based on zinc halides (24).

Shortly after the discovery of living cationic polymerization of vinyl ethers (27), the living cationic homo- and copolymerization of simple olefins by organic ester or ether BCl₃ or TiCl₄ initiating systems was demonstrated (24). The living homo- and sequential block copolymerization of isobutylene [115-11-7] and styrene [100-42-5] coinited with TiCl₄ or BCl₃ has been achieved using alkyl halide initiators in the presence of proton traps in concentrations comparable to the concentration of protic impurities in the system (16). In the absence of proton traps, however, induced chain transfer prevents living polymerization with TiCl₄. With isobutylene using BCl₃, fast polymerization by the protic impurities occurs masking the much slower living polymerization, and the monomer is essentially consumed by this process unless protic impurities are scavenged.

The key to living polymerizations is the high stability of the growing end where the nucleophilic counteranion interacts strongly with the cationic active site. Living polymerizations have also been reported with initiating systems forming nonnucleophilic counteranions in the presence of added Lewis bases (electron donors) and in the presence of common ion salts shifting the ionic dissociation equilibrium toward the nondissociated species. With these systems rapid advances have been made toward the synthesis of well-defined materials with controlled architecture, molecular weight, molecular weight distributions, and end functionalities by cationic polymerization (24).

Since the discovery of living cationic systems, cationic polymerization has progressed to a new stage where the synthesis of designed materials is now possible. The rapid advances in this field will lead to useful new polymeric materials and processes that will greatly increase the economic impact of cationic initiation.

BIBLIOGRAPHY

1.

J. P. Kennedy and E. Marechal, *Carbocationic Polymerization*, Wiley-Interscience, New York, 1982.

2.

M. Zsuga and co-workers, *Polym. Bull.* **23**, 335 (1990).

3.

J. P. Kennedy and E. G. M. Tornquist, *Polymer Chemistry of Synthetic Elastomers*, Wiley-Interscience, New York, 1968, p. 301.

4.

R. Faust and J. P. Kennedy, *J. Polym. Sci. Polym. Chem.* **A25**, 1847 (1987); **U.S. Pat. 4,910,321** (Mar. 20, 1990), R. Faust and J. P. Kennedy (to the University of Akron).

5.

R. Faust, A. Nagy, and J. P. Kennedy, *J. Macromol. Sci. Chem.* **A24**(6), 595 (1987); M. Zsuga, R. Faust, and J. P. Kennedy, *Polym. Bull.* **21**, 273 (1989).

6.

T. Higashimura and co-workers, *Macromolecules* **26**, 2670 (1993).

7.

R. Faust and J. P. Kennedy, *J. Macromol. Sci.* **A27**(6), 649 (1990).

8.

R. Faust, unpublished results, 1993.

9.

R. Faust, M. Zsuga, and J. P. Kennedy, *Polym. Bull.* **21**, 125 (1989).

10.

D. Cook, *Canad. J. Chem.* **41**, 522 (1963).

11.

J. P. Kennedy and R. T. Chou, *Polym. Preprint* **20**(2), 306 (1979).

12.

R. Faust, B. Ivan, and J. P. Kennedy, *J. Macromol. Sci.* **A28**(1), 1 (1991); L. Balogh and R. Faust, *Polym. Bull.* **28**, 367 (1992).

13.

H. C. Brown and B. Kanner, *J. Am. Chem. Soc.* **75**, 3865 (1953).

14.

J. P. Kennedy, S. C. Guhaniyogi, and L. R. Ross, *J. Macromol. Sci. Chem.* **A18**(1), 119 (1982).

15.

J. P. Kennedy and S. C. Guhaniyogi, *J. Macromol. Sci. Chem.* **A18**(1), 103 (1982).

16.

M. Gyor, H.-C. Wang, and R. Faust, *J. Macromol. Sci.* **A29**(8), 639 (1992); Zs. Fodor and co-workers, *J. Macromol. Sci.* **A30**(5), 349 (1993); M. Gyor and co-workers, *Polym. Preprint* **34**(2), 1993; Zs. Fodor and R. Faust, *J. Macromol. Sci.*, **A31**(12), 1983 (1994).

17.

J. P. Kennedy and R. A. Smith, *J. Polym. Sci., Polym. Chem. Ed.* **18**, 1523 (1980).

18.

M. Zsuga and J. P. Kennedy, *J. Polym. Sci. Polym. Chem. Ed.* **29**, 875 (1991).

19.

D. W. Grattan and P. H. Plesch, *Makromol. Chem.* **181**, 751 (1980).

20.

M. Marek, J. Pecka, and M. Maleska, *Ge Intern. Symp. of Cat. Pol.* 1983, p. 21, G. Sauvet, J. P. Vairon, and P. Sigwalt, *J. Polym. Sci. Polym. Chem. Ed.* **16**, 3047 (1978).

21.

F. R. Kalafov and co-workers, *Makromol. Chem. Rap. Commun.* **6**, 29 (1985).

22.

E. Marechal, L. Bull, and H. A. Nguyen, *Polym. Bull.* **17**, 157 (1987); L. Balogh and R. Faust, *Macromolecules*, **27**, 3453 (1994).

23.

P. Sigwalt and G. A. Olah, *Makromol. Chem.* **175**, 1077 and 1039 (1974).

24.

J. P. Kennedy and B. Ivan, *Designed Polymers by Carbocationic Macromolecular Engineering*, Hanser Publishers, Munich, Germany, 1991.

25.

A. Ledwith and D. C. Sherrington, *Polymer* **12**, 344 (1971); T. Higashimura, M. Miyamoto, and M. Sawamoto, *Macromolecules* **18**, 611 (1985).

26.

T. Biedron and co-workers, *Makromol. Chem. Macromol. Symp.* **32**, 155 (1990).

27.

M. Miyamoto, M. Sawamoto, and T. Higashimura, *Macromolecules* **17**, 265 (1984).

Rudolf Faust
University of Massachusetts, Lowell

INK-JET PRINTING.

See INKS; PRINTING PROCESSES.

INKS

According to most historians, writing ink was first prepared and used by the Chinese and the Egyptians as early as 2600 BC. These early inks were probably composed of carbonaceous materials such as lamp black or soot mixed with animal glue or vegetable oil vehicles. Reference is also made to the Chinese invention of solid ink blocks and pellets similar to India ink as it is known today; the Chinese invention had its origin during the period 220–419 AD. Writing-ink formulation became a highly developed art under the Chinese, who were printing from hand-cut blocks in the eleventh century AD, 400 years before Gutenberg introduced movable type in Europe. Writing inks differ from printing inks in that the latter are generally applied to a substrate by means of a printing press. Printing inks as supplied to the graphic arts industry are used in much greater volume by far as compared to writing inks. This article is divided into a discussion of printing inks, followed by some miscellaneous categories of ink, including ink for ball-point pens, with which the greatest amount of ink writing is done. The number of printing-ink manufacturing establishments in the United States is approximately 450. This includes some 50 captive ink plants. The value of the total printing-ink production in the United States was approximately \$2,800,000,000 in 1992 (see also PRINTING PROCESSES).

Printing Inks

Printing ink is a mixture of coloring matter dispersed or dissolved in a vehicle or carrier, which forms a fluid or paste which can be printed on a substrate and dried. The colorants used are generally pigments, toners, dyes, or combinations of these materials, which are selected to provide the desired color contrast with the background on which the ink is printed. The vehicle used acts as a carrier for the colorant during the printing operation, and in most cases also serves to bind the colorant to the substrate. Printing inks are applied in thin films on many substrates such as paper, paper board, metal sheets and metallic foil, plastic films and molded plastic articles, textiles, and glass (see FILM AND SHEETING MATERIALS, PAPER). Printing inks can be designed to have decorative, protective, or communicative functions. In some cases, combinations of these functions are achieved.

There are four principal classes of printing ink, which vary considerably in physical appearance, composition, method of application, and drying mechanism. These also fall into two general types of consistency or viscosity, paste and liquid. The classes are letter press and lithographic (litho) inks, which are called paste inks, and flexographic (flexo) and rotogravure (gravure) inks, which are called liquid inks (1).

The four key properties of inks are drying, rheology, color, and end use properties. Use properties are those considerations that determine how printed substrates function throughout all processing and usage from the time of printing throughout the useful life of the printed product.

Drying. Drying may be defined as any process that results in the transformation of a fluid-printing ink into a very high viscosity or solid film. An ink is considered dry when a print does not stick or transfer to another surface pressed into contact with it. Drying is accomplished by one or more of the following physical or chemical mechanisms: absorption, evaporation, precipitation, oxidation, polymerization, cold setting, gelation, and radiation curing.

Absorption. Some inks (eg, oil-based newspaper inks) dry by penetration or absorption into the pores of the printed stock, which has a blotter or sponge effect. This is accomplished by the gross penetration of the ink vehicle into the pores of the substrate, the partial separation of the vehicle from the pigment, and the diffusion of the vehicle throughout the paper. The ability of an ink to penetrate into paper depends on the number and size of the air spaces present in the paper, the affinity or receptivity of the stock for the ink, and the mobility of the ink.

Solvent Evaporation. Drying of ink by evaporation of solvent is generally achieved by application of heat and or air flow and depends on the volatility of the solvent(s) employed in the ink formulation. Web offset and letterpress inks which dry by solvent evaporation are usually called heatset inks. Flexographic and Rotogravure inks usually contain more volatile solvents than heatset inks and may not require as much heat to dry. In either case, a balance must be maintained between satisfactory drying of the ink on the printed substrate and adequate press stability, or resistance to drying, on the distributing rollers of the printing press during the printing process.

Precipitation. An ink may also be caused to dry by precipitation of its binder rather than by evaporation of solvent. This can be accomplished by adding a diluent, such as water in the form of steam or humidity, to a hygroscopic solvent ink system, which causes the solubility of the resin in the ink film to decrease sharply and causes it to precipitate when its tolerance for the diluent is reached. Further drying is accomplished by absorption of the solvents into the stock and then by evaporation. Another form of precipitation setting is the quick-set mechanism. This utilizes resins held in solution in a relatively poor solvent, by means of a small amount of an excellent solvent (called a sweetener) blended with it. When the ink film is printed on the paper, an amount of the solvents is absorbed reducing the content of the sweetener solvent to a point which causes the resins to precipitate and the ink to set.

Oxidation. Inks that dry by oxidation behave much like oil paint films and dry by means of the reaction of drying oils (qv) with oxygen. They contain metallic driers, which catalyze the absorption of oxygen by the drying oil (see DRIERS AND METALLIC SOAPS; PAINT).

Cold Setting. In this process, also called hot-melt, inks are applied to a substrate in a molten state and upon cooling form a dry ink film.

Gelation. Drying by gelation is accomplished by dispersing fine particles of a high molecular weight polymer, eg, poly(vinyl chloride) resin, in a latent plasticizer for the polymer. At room temperature the dispersed polymer behaves like a pigment but at elevated temperatures it solvates, causing an immediate gel formation which is dry to the touch.

Polymerization. Thermal polymerization or curing of an ink film at elevated temperatures can follow many different chemical paths. Condensation and cross-linking reactions may be accomplished with or without the use of catalysts. However, this method of drying generally has not been widely used for printing inks, except those used for metal and glass decoration, and some clear coatings.

Radiation Curing. The use of radiation curing inks has grown significantly for a number of reasons. The growing concerns about volatile solvent, the desire to conserve energy, the marketing appeal of high gloss coatings, improved resistance to abrasion and chemicals, production efficiencies, and often improved press performances have been several of the driving forces. Although there are several forms of radiant energy used in the printing industry to dry inks, only the use of ultraviolet (uv) light and electron beam (EB) energy actually cause curing of the ink or coating through a chemical change. Radiant energies such as infrared (ir) and microwave energies are also used, but these energies are used to heat the ink to drive off solvents and do not typically cause a curing (chemical change) of the ink. The use of uv and EB technologies has grown so that these curing techniques can be found in all printing methodologies: lithographic offset, letterpress, dry offset, flexographic, gravure, and screen printing.

Uv inks. The dominant system used for uv curing is based on acrylate chemistry and is cured through free-radical polymerization. The inks are composed of pigments, oligomers, reactive (acrylated) resins, monomers (often multifunctional), photoinitiators, and additives. The uv energy is absorbed by the combination of photoinitiators which generate the free radicals through cleavage, hydrogen abstraction, or energy transfer. These initiator packages are designed to utilize the maximum amount of the uv spectrum that is available as well as provide fast and thorough curing. They are highly guarded proprietary secrets.

A second type of uv curing chemistry is used, employing cationic curing as opposed to free-radical polymerization. This technology uses vinyl ethers and epoxy resins for the oligomers, reactive resins, and monomers. The initiators form Lewis acids upon absorption of the uv energy and the acid causes cationic polymerization. Although this chemistry has improved adhesion and flexibility and offers lower viscosity compared to the typical acrylate system, the cationic chemistry is very sensitive to humidity conditions and amine contamination. Both chemistries are used commercially.

EB Inks. EB inks are very similar to acrylic uv inks without the photoinitiators. The energy of the electron beam causes the acrylate double bonds to directly cleave, forming free radicals which initiate polymerization. Due to typically higher cross-linking and the absence of photoinitiators, the EB curing systems lend themselves to applications where very low odor, low off-taste, or low extraction or residual materials are important. EB systems must have an inert gas, eg, nitrogen, curing chamber since the free-radical chemistry is inhibited by oxygen. Due to this inert gas atmosphere, EB applications are usually used for web printing. Sheet systems do exist, but the mechanical difficulty of inerting usually precludes them. EB curing units are relatively costly compared to uv units, but they require much less energy to operate, are insensitive to color and ink film thickness, and produce much less heat.

Infrared and Microwave Inks. These are inks which have been formulated to absorb these radiant energies. The energy causes the inks to heat and dry through the partial evaporation of solvent. Absorption of the ink into a porous substrate can also be part of the overall drying mechanism with these inks. They have not found wide commercial success due to the variability of the ir absorption with ink color and the energy inefficiency of microwave systems in drying nonwater-based inks.

Rheology. The rheology or flow of inks is their primary physical property. In a Newtonian liquid, any stress produces a flow, and the rate of flow is proportional to the stress. But inks are generally non-Newtonian and have a nonlinear flow curve. The common terms used to describe ink rheology are viscosity (resistance to flow); yield value (stress at which a liquid starts to flow); shear thinning (decreasing viscosity with increasing agitation); and shear thickening (opposite of shear thinning). Mayonnaise is an example of a material having a high yield value but a low viscosity (2–4). Ketchup is an example of a shear thinning fluid.

Liquid Inks. The ink-distribution systems of flexo and gravure printing presses are very simple, having few inking rollers, and do not provide the means to distribute and level highly viscous inks. Therefore, the viscosity must be low, on the order of 50–150 mPa·s(cP) . Yield value must be low also to permit distribution of ink from reservoir to fountain. Excessive shear thinning should also be avoided to obtain the best printability. Low viscosity ink is preferred for fine-line flexography and shallow-cell highlights in gravure printing.

Paste Inks. Letterpress and offset inks can vary in viscosity from 500 mPa·s(cP) for a letterpress-type newspaper ink to over 100 Pa·s (1000 P) for special litho ink metal formulations. The viscous nature of letterpress and litho inks has caused press designers to use a multitude of rollers in the ink distribution unit to ensure uniform thin films and adequate transfer of ink to the printing plate. In addition to the importance of proper ink rheology as related ink formulation to roller-to-plate transfer, it also controls, in large part, the fidelity of printing, drying speed, gloss, and overprinting properties obtained on the substrate. In most cases, higher press speeds require lower viscosity inks and vice versa. Printing smooth, dense solids can best be achieved using somewhat higher viscosity ink.

Rheology is also an important color strength determinant. High pigmentation can lead to a high yield value, highly shear thinning ink, thereby forcing a compromise between color intensity and rheology.

Measurement. There are many varieties of instruments designed for measuring viscosity and other rheological properties of inks. These include capillary, orifice, rotational, and other types. Modern ink rheology measurements are made on instruments such as the Cone-Plate, Falling Rod, Rotational Torsion, and Efflux viscometers. Rotary Tackmeters such as the Inkometer and Tackoscope are also used to record tack (film splitting force) as well as ink misting and roller stability (see RHEOLOGICAL MEASUREMENTS).

Color and Coloring Materials. The third key property of all types of ink is color which may very well be the most important one to the consumer because it has such a great psychological impact. Color (qv) has three different attributes described as hue or shade, saturation or chroma, and lightness or value.

Pigments (qv) are used not only for color but also affect physical properties such as bulk, opacity, specific gravity, viscosity, yield value, and printing qualities. Different pigments of the same hue can have varying fade, heat, chemical, and bleed resistance (5,6). Judicious selection of pigment is also required, according to the use of the ink, in considering subsequent operations such as varnishing, waxing, lacquering, or laminating. Inks make use of inorganic pigments and of dyes, most commonly in the form of organic pigments, dyes rendered insoluble in one of a number of ways (see PIGMENTS, ORGANIC). Five dye families that are of interest to ink manufacturers are azo dyes (qv), triphenylmethane dyes (qv), anthraquinone dyes, vat dyes, and phthalocyanine compounds (qv) (see DYES, APPLICATION AND EVALUATION; DYES, ANTHRAQUINONE) (7,8).

Flushed Colors. Many pigments, although dry when purchased, are prepared in a water phase and are not wetted easily by nonaqueous phases. In a process known as flushing, pigment dispersions are prepared that obviate this difficulty. Phase transfer is generally accomplished by mixing the pigment in water presscake with a water-immiscible vehicle in a heavy-duty sigma blade (dough) mixer. Most of the water separates and can be removed, although complete dehydration requires further processing, usually heating under agitation in a vacuum.

Flushed colors generally result in higher gloss and color strength, and better rheology and easier use than their dry-pigment counterparts. In paste inks, they are the principal form of colorant used in the United States. In Europe, dry colors are generally used because of the greater choice of vehicles available to the formulator.

Color Concentrates. In order to provide greater versatility and, at the same time, ease of use, pigment manufacturers also offer pigments in the form of color concentrates. These are liquids or pastes with 35–65% of pigment content. Pigments in this form are very well dispersed, requiring little extra mechanical effort to achieve the desired color value or fineness of grind. The concentrates often contain surfactants and or resinous dispersants plus a solvent. They are classified according to the solvent used. Easily dispersible (ED) pigments are produced by coating freshly precipitated pigment in a slurry tank with a resinous and or monomeric dispersant and, eventually, drying and pulverizing the solids. Such pigments are also called stir-in, since they require minimum grinding effort in order to develop the full color value.

Toners are full strength undiluted pigments used to strengthen tinctorially weak batches of pigments. Occasionally, dyes are utilized as toners. The most common pigments used in ink manufacture are as follows.

Black Pigments. The only black pigment used to an appreciable extent in inks is carbon black. It is used in newsprinting, publication, commercial and packaging printing; therefore, in large quantities. Black pigments are offered in fluffy or beaded forms and in a variety of particle sizes and physical properties.

White Pigments. Opaque white pigments commonly used in inks, in order of decreasing opacity, are titanium dioxide and zinc oxide. TiO₂ is by far the most popular white pigment. Mixtures of whites are often made with the various colored pigments to add opacity or to make pastel colors.

Transparent white pigments (extenders) commonly used in inks, in order of decreasing transparency, are alumina hydrate, magnesium carbonate, calcium carbonate, blanc fixe (precipitated barium sulfate), talc, and clay. Extenders are sometimes used to reduce the color strength and change the rheology of inks.

Inorganic Color Pigments. Iron blue is made in several shades, such as Milori blue and Prussian blue. It is economical, but has poor alkali resistance. It is also used as a toner in some black inks.

Organic Color Pigments. Printing inks are the largest consumer of organic color pigments of any industry. Yellow pigments consist mainly of diarylide yellows, Hansa yellows, and yellow lakes. Lakes are produced from dyes by depositing them on alumina hydrate. Hansa yellows are very strong and permanent pigments. They are resistant to many chemicals and are produced from azo dyes (qv) in a number of hues. Diarylide yellows, the most widely used yellow pigments, are very strong, usually not as fast to light as Hansa yellows, but more transparent. Orange pigments include dianisidine orange, diarylide orange, and Persian orange lakes.

Red Pigments. There are more red pigments used than any other organic type and they range in shade from yellowish red to deep maroon and from dirty reds to very clean, brilliant reds. A few of the more common types are as follows. (1) The para reds and toluidines are fairly fast to light and semitransparent. They are used in poster and label inks. (2) The lithol reds are available in a wide range of shades with moderate permanency to light. (3) Lithol Rubine reds are widely used as magentas in all types of four-color process printing. (4) Rhodamine reds are brilliant and have good lightfastness, when laked with phosphotungstic (PTA) and or phosphomolybdic (PMA) acids. They are extremely costly, and for this reason, are not widely used as process magentas. (5) Red Lake C is used for making the clean, brilliant orange or warm red shades. (6) Watchung reds or Permanent Red 2B pigments are

fairly clean and have fair bleed and product resistance. (7) Naphthol reds are relatively permanent and are soap-resistant. They are available in yellow and blue shades.

Blue Pigments. Phthalocyanine blues are the most widely used organic blues. They are available in both greenish and reddish shades. They have excellent lightfastness, excellent resistance to most chemicals, and are used as cyans in all four-color process printing. Victoria blue is a clean, red-shade royal blue. It is only moderately fast to light and finds most of its applications in packaging printing. Alkali blues of deep reddish shades find extensive use in all types of inks, often for toning black inks. Alkali blue toners are also supplied in flushed form.

Purple Pigments. Methyl violet is the most commonly used purple pigment. It is also widely used for toning black inks. In packaging, although very expensive, Carbazole and Vat violets are used where permanence and resistance properties are needed.

Green Pigments. Phthalocyanine green, the most widely used type, is a permanent and clean pigment. Some other organic greens, such as Malachite green, are used in packaging.

Daylight Fluorescent Pigments. These are used in all inks where brilliant fluorescent shades are required. The main applications are in packaging, greeting cards, wrapping paper, and posters. They generally are not very fast to light since they consist of dyed resins which are then pulverized to resemble a pigment.

Metallic Powders. These are usually either aluminum or bronze flakes and vary in shades from silver to gold, depending on the composition of the metal used. The silver powders can also be toned with organic pigments to produce golds or copper shades using transparent yellow or red pigments.

Other Ingredients.

Driers. These are generally soaps of cobalt, manganese, and other metals formed with organic acids such as linoleic, naphthenic, and other organic acids. They catalyze oxidation of drying oils (qv), and thus are used in inks that dry by oxidation (see **DRIERS AND METALLIC SOAPS**).

Waxes. These are dispersions of polyethylenes, Fischer-Tropsh, Teflon, or vegetable waxes (qv) in a vehicle or solvent. They impart slip and scuff resistance to ink films. Polyolefin waxes and Teflon are also available as powders that can be directly mixed into inks.

Antioxidants. Widely used antioxidants are eugenol, ionol, and the like. They retard premature oxidation of inks on the press rollers when used at low concentrations.

Miscellaneous Additives. These include lubricants, surfactants, thickeners, gellants, defoamers, and preservatives.

Letterpress and Litho Newsprint Inks

The U.S. news ink industry, with a sales potential of over \$300 million annually, represents a dynamic, ever-changing segment of the graphic arts market. Forces changing this industry stem from government regulations, shifts in market needs, recurring shortages in petroleum-based raw materials, and novel technologies developed through R & D effort. Concurrent to these, the introduction of faster printing presses and new printing modes such as publication flexo further influence the evolution and change in ink technology.

A gamut of products available commercially meet the needs of technical specifications, environmental trends, and quality requirements. From the three printing modes currently available, the web offset lithographic process and water-based flexo continue to grow at the expense of letterpress printing. The change in the market share is governed by the growing demand for print quality, which the letterpress process cannot deliver.

The printing of newspapers is conducted at very high speeds, often reaching 3000 feet per minute. All three printing processes utilize similar quality newsprint which, essentially, is made of groundwood or thermomechanical pulp. Presses are fed a continuous web of newsprint that unwinds from a feed roller. Inks dry by absorption of liquid into the porosity of the substrate. Some evaporation of water in a flexo publication ink can accelerate the drying process.

Web Offset. This is, by far, the largest type of newspaper printing. Three different press ink feed configurations used by the industry, ie, open fountain, injector, and keyless, require inks with specific rheological characteristics. Web offset lithographic printing uses planographic, aluminum-based printing plates, fountain solution, and an ink formulated to accept and emulsify water.

Lithographic inks are printed from planographic plates on which the image is neither raised nor depressed. The image is differentiated by a chemical difference from the nonimage area. In modern aluminum offset plates, the nonimage areas are anodized aluminum which has been chemically or mechanically roughened (grained). The image area is formed photographically or by direct laser imaging using a photopolymerizable coating. The unexposed areas which form the nonimage are washed off using an aqueous developing liquid, exposing the anodized aluminum. An aqueous solution of salts, buffers, colloidal gums, surfactants, and other materials, known as a fountain solution, is used to prewet the nonimage areas of the plate and make them hydrophilic. The ink is then rolled over the entire plate, but since it is highly oleophilic it only adheres to the image areas which have been formed by the photopolymerized hydrophobic coating. This image is then transferred via a rubber blanket cylinder directly to the printing substrate or paper. The plates are relatively inexpensive and easily made, which is one of the big advantages of lithography. The ink must be able to emulsify some quantity of the fountain solution in a stable manner in order to have the ink transfer properly during this process. The interactions between the ink and the fountain solution thus become the critical factor in lithographic printing. It is important that proper formulation of the ink vehicle and fountain solution be done with a view toward controlling the surface chemistry of both ink and fountain solution, otherwise ink tends to deposit in the nonimage areas of the plate, causing scumming. The inks are formulated with resins and varnishes as well as surface-active additives to facilitate this emulsification process. Despite this apparent complexity, this process can yield excellent print quality when all three elements, ie, plates, fountain solution, and inks, are properly formulated and the printing parameters are maintained at optimum levels.

Inks. Web offset inks are highly pigmented, to yield the desired print density at a thin printing film (~1 μm). The viscosity of offset inks is relatively high, but varies with the press configuration. For example, low and high shear viscosities are the highest for open fountain inks, and lower for injector and keyless inks. The flow characteristics for these inks are in reversed order; thus keyless inks possess the most flow whereas open fountain inks the least. Inks of two distinct chemistries are used in this process: a traditional type based on mineral oil, and a newer one containing soya bean oil. The resins employed are selected from low cost nonfunctional hydrocarbon-type resins, more complex hydrocarbon resins (qv) modified with rosin and or phenolics, and gilsonite (for black only).

Although the black inks are predominantly based on mineral oil, colors are almost entirely formulated with a soya bean oil vehicle. The superior printability of colors and economics of blacks guide the selection of product types by the market. Recently developed low rub blacks offer smudge-resistant print. Their share of the market is growing rapidly. The low rub characteristics of these inks are produced through the use of low structure carbon black. The addition of resin further enhances the smudge resistance but imposes a premium price.

A variety of additives are used to control the properties of wetting and dispersion of pigments, flow, lithography, and rub-off of inks. These additives belong to classes of materials such as surfactants, bentonite clays, alkyds, functional resins, polymers, etc.

Letterpress. This is the oldest printing process still in use. It continues to be replaced by newer printing processes. Printing is conducted from a raised image area of the printing plate. Inks in the printing process are transferred directly from a raised area to a substrate. The printing plates contain a thick layer of photopolymer (often a mixture with polymer such as poly(vinyl alcohol) deposited over a plastic or aluminum base.

Inks. Basic raw materials for letterpress inks, such as mineral oils, soya bean oil, resins, and pigments, are essentially the same as those used in web offset inks. Inks are tinctorially weak, relatively fluid, and their low and high shear viscosities are low.

Formulas are very simple. For example, in the case of black they often contain no additives and consist merely of pigment, mineral oil, and asphaltic pitch. Low rub inks are available; however, due to economics, a traditional type of formulation based on mineral oil and high structure carbon black is predominantly used.

Flexo. The use of water-based flexographic ink for newspaper printing was initiated in the late 1970s. In comparison to letterpress or web offset printing, this is a relatively young and upcoming printing process. The exclusion of oils from an ink vehicle lead to the elimination of undesirable strikethrough, with resulting brilliant colors which can be observed in the Sunday issue of the comics section. Despite its success in some areas of the market, this process still represents a small segment of the newspaper market (6–7% in 1992).

Printing is conducted with a printing plate similar to letterpress. However, the chemistry of the photopolymer is somewhat different in order to

make the plate water insensitive. A high quality of print has been demonstrated by several newspapers utilizing this printing process. The printed matter is virtually smudge-free.

Inks. Typical inks are water-based, with acrylic emulsion resins as the main binder. Inks of this type occasionally use natural products such as starches, lignins, and lignin derivatives. Hence, ecologically, this process is more desirable. Practically all resins used in this ink system are rendered water soluble or emulsifiable through neutralization with organic amines. Strong absorption of amine by the newsprint renders the resin, after printing, water insoluble and prevents bleeding of printed matter. Press ready inks are very fluid and of low viscosity. Inks contain a variety of additives for the elimination of foaming, dispersion of pigments, rheological modifiers, slip agents, etc.

Web Heat-Set Publication and Commercial Inks. Almost all heat-set inks are now printed on web offset presses, and are based on vehicles containing synthetic resins and or some natural resins. These are dissolved in hydrocarbon solvent fractions which are specially fractionated for use in the ink industry. They vary in boiling range between 180 and 300 °C. Small percentages of alkyd resins (qv) may be contained in these inks. They dry in less than one second by means of solvent evaporation in a heatset oven. These ovens utilize high velocity hot air to raise the web temperature to 120–150 °C.

Sheet-Fed Offset Inks. A large segment of commercial printing is done on sheet-fed presses almost entirely by the litho process. Inks for these presses are based on vehicles containing phenolic modified, maleic-modified, or unmodified rosin-ester resins dissolved in vegetable drying oils and diluted with hydrocarbon solvents. Some inks also contain alkyds, which may be modified with other polymers, such as urethanes, styrene, and the like. On coated stocks, many sheet-fed inks quick-set to a tack-free state by precipitation and solvent separation and then dry fully by oxidation. Air (oxygen) upon reacting with unsaturated oils forms a highly polymerized, tough ink film. The most commonly used oils are linseed, soya, and tall oils. Special acrylic resins have been developed for use in quickset inks, and offer nonskinning properties and excellent press stability.

Duplicator and Business Form Inks. Duplicator sheet-fed machines require very press-stable yet quicksetting inks. They must also possess good lithographic properties of wide tolerance for fountain solution and provide good printing properties on a wide variety of uncoated papers. The inks can contain drying oil alkyds along with hydrocarbon resins and high boiling (200–370 °C) hydrocarbon solvents. Business form inks closely resemble the lithographic heatset or quickset inks. Business forms have also been printed by the ink jet method. These inks are usually based on water, glycols, and dyes.

Folding-Carton Inks. The majority of folding-carton inks are based on various quickset vehicles, as described above. However, when maximum gloss, good rub, and product resistance are required, they contain mainly oleoresinous vehicles. These vehicles are composed of phenolic, phenolic-modified, and maleic-modified and or esterified rosin resins which have been dissolved in drying oils or alkyds or both. They dry by oxidation to form tough, glossy films. Ultraviolet light-cured inks are also used in litho printing of high quality folding cartons.

Metal Container Inks. Ink vehicles for metal containers that are printed on special flat sheet-fed litho presses are based mainly on blends of oleoresinous varnishes containing alkyds, polyesters, and melamine resins. These inks dry during a 5–15 minute cure at 150–250 °C in long gas-fired ovens. Polymerization, oxidation, and cross-linking reactions accounts for their drying and hardening. Ultraviolet inks also have been used in metal decorating, but are not popular in this type of printing.

Preformed Two-Piece Metal Containers. Ink vehicles for letterset printing of two-piece aluminum or steel containers are mainly based on special polyester vehicles used in conjunction with melamine cross-linkers. Short cycle ovens which dry inks in 1–5 seconds are now used and operate at temperatures as high as 350 °C. The rheology of these inks must be adjusted to the unique geometry of the press. Desired rheological properties are achieved by the use of additives as well as extender pigments.

Plastics. Vehicles in offset inks for plastics (polyethylene, polystyrene, vinyl) are based on hard drying oleoresinous varnishes which sometimes are diluted with hydrocarbon solvents. Letterset inks for polystyrene employ vehicles of somewhat more polar nature. Polyester or other synthetic resins (acrylic) dissolved in glycol ethers and or esters are used in some of the older inks. Uv inks are widely used for decoration of these preformed plastic containers.

Manufacture. Paste inks are produced in two ways: (1) by mixing predispersed (preground) or flushed pigment concentrates with vehicles, solvents, oils, and compounds, and filtering or (2) by mixing dry pigments or resin coated pigments with vehicles and compounds and then dispersing them with various types of ink mills. Mixing is done in change-can mixers, or in large agitated tanks which can hold up to 10,000 kg. A variety of equipment is being employed, from simple agitators to high speed, sawtooth blade mixers or concentric, double-blade mixers in order to achieve the desired predispersion. Milling can be done on three roller horizontal or vertical mills and in media mills. The three-roller mill is still used for grinding small batches of viscous paste ink. Its steel rollers are polished and accurately machined to form a crown in the center. A speed differential exists in rotation of each of the rollers with respect to each other. They also can be set to provide greater or lesser clearance between them. The dispersion of the pigment is accomplished by shearing forces generated by this differential speed as well as by the closeness of roller setting. The shot mills break pigment particles by crushing action of the speeding shots. The smaller the size of the shot and the higher loading and the more complex (tortuous) path in the chamber, the better the dispersion. Inks may be heated so that they are low enough in viscosity to be ground in media mills (vertical and horizontal). Finished inks are packed in metal cans (0.5–5 kg), in metal or plastic pails (8, 11, and 19 L), or in 114- or 190-L metal or fiber drums, and for volume users, in large tote bins. The more fluid inks (news, flexo, or gravure) usually are delivered in tank trucks directly to the printer. Ink vehicles are usually produced in separate resin varnish plants. Here 1,900–19,000-L reactor kettles are equipped with powerful agitators, rellux condensers, liquid traps, gas and liquid pumps, and ducts for material charge. Heating is done by multistage Dowtherm or hot-oil jackets, or in some installations, electric heaters. Temperatures and reaction rates are automatically monitored and controlled. Finished vehicles are filtered routinely via bag or cartridge filters. Synthesized resins can be cast or chipped in solid form from the reactor kettles and also converted directly into fluid varnishes in a single operation.

Quality Control and Testing. Control of inks is done by examining their color strength, hue, tack, rheology, drying rate, stability, and product resistance. Elaborate control equipment and laboratory testing procedures are employed to test the finished inks. Weather-Ometers, Fade-Ometers, glossmeters, printability testers, colorimeters, spectrophotometers, viscometers, rub testers, and gas chromatographs are employed to check production batches or to pretest new submissions or raw materials. Proofing presses and sometimes pilot presses are utilized by ink manufacturers to control production and test new formulations. The move toward higher printing speeds and quality, and greater economy and reduced volatile organic compounds (VOCs), necessitates intensive ink research and development efforts. New regulations, availability and cost of raw materials, new or modified substrates, and faster presses require constant updating of the formulations. Considerable research and development time and expenditures have been devoted to low VOCs and low temperature drying inks. Ultraviolet, electron beam, and water-based inks for various applications are some of the areas of research and development being actively pursued. Trends to a more universal use of renewable resources have led to increased amounts of natural products such as vegetable oils and rosin in litho inks of all types, particularly in sheet-fed litho inks.

Flexographic and Rotogravure Inks

Flexo and gravure inks are both known as liquid inks because of their low viscosity. The inks for both systems have basic components in common with inks for other printing processes. Vehicles disperse and carry the pigment, and also contribute most to the end use properties. Colorants provide color. Solvents dissolve resins in the vehicle and determine drying rate. Additives modify ink properties to overcome deficiencies.

The vehicle is composed of resins, solvent, and additives. The resins impart adhesion and end use resistance properties to the film. Resins are polymers which can be film formers or nonfilm formers (Table 1). Film formers are flexible and form a continuous film when dry. Some resins require the use of an additive such as a plasticizer to achieve film formation. Nonfilm formers are brittle polymers which do not form a film even with plasticizers (qv). Plasticizers are nonvolatile liquids or soft resins which may partially dissolve the main resins. Resins for flexo ink must be soluble in solvents which will not harm the printing plate or the substrate. Resins for gravure ink do not have this handicap because of the metal gravure plate cylinder. Colorants are pigments or dyes. Pigments are chemical compounds which are insoluble in either the solvent or the resin. Because of this insolubility pigments must be mechanically dispersed into a vehicle and broken down to small particle size. The dispersion and grinding process is required to develop the color strength and keep the pigment stable in the vehicle. Dyes are chemical compounds which are soluble in either the solvent or the vehicle. Dyes give strong, brilliant colors compared to pigments but normally have poor resistance properties. Solvents are required for two reasons. The first requirement a solvent must satisfy is to dissolve the resin; this results in a low viscosity ink suitable for printing. Secondly, the solvent must evaporate quickly and completely from the printed film.

Table 1. Resins for Flexographic and Rotogravure Ink Vehicles

Resi	Flexo	Gravure	Film former
acrylic	†	†	yes
cellulosics	†	†	yes
chlorinated rubber		†	yes
epoxy	†	†	yes
nitrocellulose	†	†	yes
polyamide	†	†	yes
polyester	†	†	yes
polyketone	†	†	
polystyrene		†	yes
polyurethane	†	†	yes
rosin-based	†	†	
Saran		†	yes
shellac	†		yes
vinyl		†	yes

Both flexo and gravure inks are delivered in the form of a virgin ink concentrate, which retards the speed of pigment settling and reduces shipping costs. Solvent is used press-side to reduce the ink to a correct printing viscosity. One must choose the solvent carefully. A solvent which evaporates too quickly may dry on the flexo plate or plug up the gravure cylinder cells. A solvent which evaporates too slowly results in too much retained solvent in the ink film, possibly causing odor, blocking, or resistance problems. In flexo and gravure printing, too slow an evaporation rate results in poor trapping of each subsequent wet ink. A commonly overlooked difficulty when dealing with solvent combinations is the relative evaporation rates of the two solvents. A solvent ratio in the ink of 50 :50 does not necessarily evaporate from the ink fountain in the same ratio. If solvent lost by evaporation is replaced in the original ratio of 50 :50, the solvent combination may be too strong, resulting in retained solvent in the film or drying problems.

Additives are used to provide a specific property. For example, a wax provides rub resistance in the printed film or a surfactant reduces foam generation in the fountain.

Manufacture. Manufacturing processes consist of two general operations, vehicle preparation and pigment dispersion. Vehicle preparation can be as basic as polymerization of resins or as simple as cold dissolving of vehicle solids in appropriate solvents. Therefore, vehicle preparation equipment includes autoclaves for polymerization reactions and high speed mixers for simple dissolving. Pigment dispersion can be done in a ball mill which lends itself to volatile fluid formulations, or in a vertical or horizontal media mill. Pebble mills using porcelain balls and linings are used for white and light colors because steel ball mills, although more effective, cause discoloration. Darker colors such as reds, blues, and blacks can be made in steel ball mills. Much of the pigment dispersion done in the 1990s is in a horizontal or vertical media or shot mill.

A premix of pigment and varnish is forced through a cylindrical cavity filled with mechanically agitated shot or sand. The grinding takes place as the pigment clumps are forced through the small openings between the moving shot. Color concentrates are made in dough mixers and pigment resin dispersions, called chips, on two-roll rubber mills. These high shear methods often result in better dispersion and consequently higher gloss than is achieved with ball or media mills.

Rubber-mill chips are dissolved similarly to resins, to provide color concentrates. Dough mixer and chip concentrates must be diluted with solvent and other vehicles to make finished inks. Media milling is becoming a method of choice in both flexo and gravure ink manufacturing. Other high speed dispersing units, such as the Morehouse, Cowles, Kady, and others, are also used.

Rotogravure Inks. Since there are no rubber or plastic components in contact with the solvents contained in gravure ink formulations, it is permissible to use solvents such as ketones and aromatic hydrocarbons which cannot be tolerated in flexo inks. This provides the gravure ink formulator with much greater latitude in regard to binder selection. In other respects the compositions generally are similar.

Ink Types. There are 10 gravure ink types categorized by the binders or solvents used: A, aliphatic hydrocarbon; B, aromatic hydrocarbon; C, nitrocellulose; D, polyamide resins; E, SS nitrocellulose; M, polystyrene; T, chlorinated rubber; V, vinyls; W, water-based; and X, miscellaneous.

Solvents. Common terminologies used interchangeably are solvents, diluents, reducers, and thinners (Table 2). Technically, solvents are materials that completely dissolve resins in the ink vehicle. Diluents are liquids that may not completely dissolve the resin by itself. Solvents can also be thinners, but most often thinners are blends of solvents and diluents. Reducer is another name for thinner, referring to the solvent blends used to reduce the viscosity of a virgin ink on the press to running viscosity.

Table 2. Commercially Used Flexo and Gravure Solvents^a

Solvent	Comparative drying time ^b	Boiling range, °C	Flash point, °C	Density at 20°C, g / L
<i>Alcohols</i>				
methyl alcohol	10	64–65	16	791
ethyl alcohol, 99° o anhydrous	18	75–80	18	791
ethyl alcohol, 95° o	20	75–80	18	815
isopropyl alcohol, 99° o	27	81–83	21	785
<i>n</i> -propyl alcohol	55	95–98	29	803
<i>sec</i> -butyl alcohol	63	99–100	29	801
isobutyl alcohol	77	106–109	39	801
<i>n</i> -butyl alcohol	125	116–119	46	809
<i>Aliphatic naphthas</i> ^a				
hexane	7	66–70	–18 ^c	683
fast diluent naphtha	7	60–82	–18 ^c	689
heptane (tolu-sol)	15	93–104	–18 ^c	725
lacquer diluent	17	93–116	–1 ^c	743
octane	31	102–110	–1	743
VM&P (varnish makers' and painters')	51	121–149	10	755
naphtha				
mineral spirits	600	154–204	38	779
<i>Aromatic hydrocarbons</i> ^d				
toluene	26	109–112	7	868
xylene	88	135–143	27	870

		<i>Esters^e</i>		
methyl acetate, 80° o	5	53–59	–1 ^c	899
ethyl acetate, 88° o	10	72–80	6	887
isopropyl acetate	12	84–90	16	869
<i>n</i> -propyl acetate	22	95–103	18	881
<i>sec</i> -butyl acetate	33	104–117	32	860
isobutyl acetate	36	110–119	31	863
<i>n</i> -butyl acetate	63	115–130	39	875
amyl acetate	100	120–150	47	863
Cellosolve acetate	330	145–165	49	971
		<i>Glycol ethers</i>		
Dowanol PM	88	118–126	38	917
methyl Cellosolve	130	123–126	46	962
Cellosolve	190	132–137	54	928
butyl Cellosolve	1000	166–173	74	804
		<i>Ketones^f</i>		
acetone	5	56–57	–9 ^c	791
methyl ethyl ketone	11	78–81	–1 ^c	804
methyl isobutyl ketone	37	114–117	24	801
methyl isoamyl ketone	150	143–146	43	813
cyclohexanone	270	130–173	54	944
		<i>Nitroparaffin</i>		
2-nitropropane	50	119–120	39	987

^a Acceptable for flexo if used with buna rubber plates and rollers.

^b The drying-time comparisons are based on the use of ethyl acetate as standard at 10. Acetone is about twice as fast as ethyl acetate and is listed as 5.

^c At very low temperatures, flash-point determinations are inconsistent; flash occurs below the temperature shown.

^d Cannot be used in flexo.

^e Should not exceed 25° o of flexo solvent.

^f Flexo usage restricted to butyl rubber plates and rollers.

Solvents for A-type inks are aliphatic hydrocarbons, for example, hexane, textile spirits, Apco Thinner, lactane, VM & P (varnish makers' and painters') naphtha, and mineral spirits. Aromatic hydrocarbons such as toluene and xylene are solvents for B-type inks. Generally, a blend of aliphatic and aromatic hydrocarbons is commonly used for this type of ink.

Ketones and esters are required for C-type inks. Types of esters are ethyl acetate, isopropyl acetate, normal propyl acetate, and butyl acetate. From the ketone class, acetone or methyl ethyl ketone (MEK) can be used. The usual solvent for D-type inks are mixtures of an alcohol, such as ethyl alcohol or isopropyl alcohol, with either aliphatic or aromatic hydrocarbons. Commonly used mixtures are 50 50 blends by volume of alcohol and aliphatic hydrocarbon.

The alcohols, proprietary denatured ethyl alcohol and isopropyl alcohol, are commonly used for E-type inks. Many E-type inks benefit from the addition of small amounts of ethyl acetate, MEK, or normal propyl acetate to the solvent blends. Aromatic hydrocarbon solvents are used for M-type inks. Polystyrene resins are used to reduce the cost of top lacquers. T-type inks are also reduced with aromatic hydrocarbons. Acrylic resins are used to achieve specific properties for V-type inks. Vehicles containing vinyl chloride and vinyl acetate copolymer resins make up the vinyl ink category. Ketones are commonly used solvents for these inks.

W-type inks use water, or mixtures of water and alcohol, as the solvent. Inks which are not of a recognized type are classified as X-type. The solvent required is specific to the ink formula and the ink maker makes proper recommendations.

Water-Based Inks. Approximately 50° o of all flexographic inks use water as their primary solvent and diluent. They contain vehicles based on either acrylic emulsions, or hydrosols or an alkali-soluble rosin ester having a high acid number such as partially esterified fumurated rosin and shellac. Carboxylated acrylic polymers, usually containing some styrene, have largely replaced natural resins because they provide better abrasion and water resistance. Ammonia or other volatile amines are used to solubilize these carboxylated resins and form resin salts. The volatile alkali evaporates from the ink film, rendering the printed matter water resistant.

Main advantages of water inks include excellent press stability, printing quality, heat resistance, absence of fire hazard, and the convenience and economy of water for reduction and wash-up.

Applications of Gravure Inks. The majority of Type A and Type B inks are used for gift wraps, newspaper supplements, catalogues, advertising inserts, and similar publication work. Inks in the Type C group are used for printing on foil, paper, cellophane, paperboard, coated and uncoated paper, glassine, acetate, metallized paper, and some specialized fabrics. Type C inks are the dominant group used in packaging gravure. Type D inks have excellent adhesion to many plastic films. They are used in foil, paper, and paperboard as well as on a variety of films. Type E inks include a wide variety of inks and lacquers and some dye inks. They are often used on paper and paperboard, some grades of cellophane shellac, or nitrocellulose primed foil pouch stock glassine and many specialty coated papers and boards.

Extremely good resistance to alcohol and soap are two of the unique characteristics of Type T inks. They are regarded as high quality inks exhibiting high gloss excellent printability and heat resistance. Use of Type W gravure inks, primers, and lacquers has been motivated by the need to comply with VOC emission standards. Water inks are primarily used in packaging gravure on board and paper. Publication gravure printers are actively testing Type W inks for various publication applications. Type V inks are used for printing vinyl films and Saran.

Lamination Inks. This class of ink is a specialized group. In addition to conforming to the constraints described for flexo and gravure inks, these inks must not interfere with the bond formed when two or more films, eg, polypropylene and polyethylene, are joined with the use of an adhesive in order to obtain a structure that provides resistance properties not found in a single film. Laminations are commonly used for food applications such as candy and food wrappers. Resins used to make this type of ink cannot, therefore, exhibit any tendency to retain solvent vapor after the print has dried. Residual solvent would contaminate the packaged product making the product unsalable.

Even though high molecular weight polyamides are not nitrocellulose compatible (thereby eliminating the possible use of the commonly used nitrocellulose color bases) they do meet the requirements described previously and form the basis for most solvent-based laminating ink systems. High molecular weight polyamides disperse pigments well. A single color base line specifically for use in laminating ink can therefore be prepared.

In addition to polyamide, lamination inks ordinarily contain modifiers such as polyketone resin, plasticizer, and wax to impart specific properties such as block resistance and increased bond strength. Because laminating inks are usually reverse-side printed and end-up sandwiched between films, gloss is not a primary requirement. Water-base laminating inks that will meet the U.S. EPA emission requirements and have the correct functional properties are currently under development.

Miscellaneous Inks

Screen Process Inks. These inks, often known in the past as silk-screen inks, are printed on the substrate by being forced through a screen stencil by means of a squeegee. For many years this was a hand-printing operation, but it has now become largely mechanized. Screen-process inks are dispersions of pigments in vehicles which are, for the most part, solutions of resins in solvents of the boiling range of VM & P naphtha. Drying of solvent-based inks is usually by evaporation, but in some cases it is a combination of oxidation and evaporation. Various types of binders are used such as rosin esters, phenolics, cellulose derivatives, vinyls, and oleoresinous varnishes, depending on the film properties desired. Uv inks are also widely used for screen-process printing. After premixing, the ink is ground on a three-roll or media mill. The resulting ink should be short and soft so as not to drag on the squeegee and to release the substrate cleanly after the print is made.

The screen-process method of printing can apply a thicker film of ink to the substrate than other printing processes, and for this reason is particularly adapted to applications where maximum opacity is desired, or where fluorescent pigments are used, or for printing etch-resists for printed circuits. Screen-process printing also is used for printing of both board and paper posters, metal signs, glass, ceramics, and plastics. The substrate need not be flat; the process also accommodates round or oval objects.

Stamp-Pad Inks. These inks are impregnated into a cloth or foam rubber pad and transferred by pressure to rubber type which is then stamped or impressed against the substrate. The inks must be completely nondrying in the pad and yet dry by rapid penetration into the paper. Since it is desirable that the total ink soak into the stock, dyes are used rather than pigments. The vehicles used are usually glycols.

Ball-Point Inks. These inks are medium-viscosity semi-Newtonian fluids of high tinctorial strength which must be slow drying and free of particles so that they continue to feed to the paper without clogging. Drying on the paper is accomplished by rapid penetration and some evaporation. These properties are obtained by strong dye solutions and pigment dispersions in vehicles containing oleic acid and castor oil or a sulfonamide plasticizer. Rheology of these inks exhibits modest thixotropy which prevents their leakage through the openings around the ball.

Water-Based Writing Inks. These consist of very fine pigment dispersions in aqueous media containing small amounts of glycol or glycerol and a dispersing aid. They dry mainly by evaporation and quick wetting of cellulosic fibers in paper substrates.

Engraving Inks. Steel-plate printing is an intaglio process, or copper in which the image is etched or engraved in continuous nonscreened lines on a steel plate. The ink is applied in a heavy layer to fill the engraving and then wiped off the nonprinting surface with paper or a rotating plastic roller, leaving the ink only in the engraving, which is then pressed against the paper with very high pressure to deposit the ink on the paper. This process is used for high quality stationery, stock certificates, and paper currency. Owing to the thick film that can be deposited (ca 25–150 μm), high strength formulations are not required, but the body of the ink is quite short so as to wipe cleanly from the plate. Drying is by a combination of oxidation or polymerization and by evaporation of solvent. The pigment, including a large percentage of colorless extender pigment, is dispersed on a three-roll mill in a vehicle composed of heat-bodied drying oil or oleoresinous vehicle, sometimes in combination with a resin–solvent type vehicle. Web engraving presses using heat or electron beam curing have been developed. They use appropriate polymerizing vehicles. Owing to the thick film of ink deposited, sufficient flexibility must be present to prevent brittleness on aging.

Electrostatic Inks. Electrostatic printing is accomplished by causing charged colored particles to move in an electrostatic field to a substrate in the form of an image. The image may be formed by a screen stencil or by a gravure cylinder. One of the primary advantages of this process is its ability to print across gaps and thus without pressure. The electrostatic ink, also called an electrostatic toner, is a powder composed of pigment dispersed in a resin. The particles must have the proper electrical properties, particle-size range, and be free-flowing. After the image is deposited on the substrate, it is heat- or solvent-fused to a continuous film. Pigments and resins must be chosen to meet the application requirements and at the same time to satisfy the physical and chemical resistance requirements of the process (see ELECTROPHOTOGRAPHY).

Decal Inks. Decalomania is a transfer method of printing. The design is first printed on a temporary base by lithography, letterpress, gravure, or screen process, depending on the detail and thickness of image desired. Usually the temporary base is paper which has been coated with a water-soluble material. The inks must dry completely on the surface by oxidation or solvent evaporation because the treated paper has no ink absorbency. After the initial printing, the design is transferred to the permanent substrate by direct contact and soaking with water. The formulation of decalomania inks is governed by the particular printing process employed in printing the transfer paper. Decalomanias for ceramics require pigments that may be heated to high temperatures. Further, most decalomanias should use pigments that are fast to light because many are subsequently transferred to outdoor signs or to store windows. Vehicles consist of oleoresinous varnishes containing metallic driers or are resin–solvent types.

Hot-Transfer Inks. Hot-transfer printing is similar to the decalomania process in that the printing is first done on a temporary substrate, but heat is used as the transfer mechanism rather than water solubility. One type of hot-transfer ink is made with heat-fusible resins and waxes to be transferred to cloth. These inks should penetrate into the cloth and not be affected by subsequent washing of the fabric. In the packaging industry, labels are printed on a web of special coated paper by conventional printing such as gravure. This web is then fed to automatic labeling equipment. The relationship between the paper web and the ink is such that the ink is immediately released by heat and transferred to the surface of the package, such as a plastic bottle, to which it then adheres as a permanent label.

Ink-Jet Printing. This is a noncontact form of printing that has the advantage of including all the information on a printed page in digitized form in computer memory, thereby eliminating the need for a plate. One principle of operation involves the issuance of liquid ink from an orifice at very high speed to form a jet which is then broken up by ultrasonic energy to produce uniform droplets that can be charged electrically. These droplets can be deflected electrostatically into a catcher, while the uncharged droplets continue in flight to form dots on the printing surface to construct images. Another principle of operation is the drop-on-demand jet which emits droplets of ink only when energized by the computer. Ink-jet technology provides a means of fast, dependable, quality, single or personalized copy printing. Its nonimpact nature permits printing on uneven surfaces and delicate materials. Computer operation permits the encoding of both repetitive and nonrepetitive information.

The inks formulated for jet printing must be very fluid, stable, and free of any particles that could cause clogging of the jet nozzles, and be capable of depositing and adhering to a substrate with a minimum of character fogging. They are generally formulated with soluble dye colorants in a suitable aqueous or solvent-based vehicle (9).

Environmental Considerations

General environmental concerns such as use of renewable resources rather than crude-oil-based chemistries, biodegradable inks and coatings, the pressure to recycle waste materials back into the raw material supply, etc, impact printing ink technology. Technologies developed in response to environmental concerns have been emerging and will continue to emerge as new issues supplant existing ones.

The United States has been among the most highly regulated countries in the world. There has been a persistent flow of legislation, governing not only virtually all printing processes but the use and disposal of printed matter as well. A very clear, historical account of the effect of antipollution laws on all aspects of printing can be traced. These laws have indirectly driven many changes in printing ink technology in order to accommodate the concerns of the primary customers, the printers.

Since the late 1980s, similar regulatory forces acting upon the Canadian and European markets have been seen. However, the heightened perception of environmental problems is propelling these countries at a faster rate than has been the history in the United States. As printing ink supply becomes a global business, the successful marketing of ink technology will require attention to an expanding body of non-United States regulations.

United States regulations encompass both federal, state, and local guidelines. In addition, there are numerous voluntary industry guidelines affecting ink making.

Occupational Safety and Health Act. OSHA regulations deal principally with physical aspects of safety and those things generally associated with accident prevention. These federal regulations deal especially with the need for established material safety data sheets and the proper labeling of printing inks under the Hazard Communication Laws.

State Right to Know Laws. These state guidelines can be thought of as an extension of the federal OSHA requirements and primarily affect the labeling of printing inks in various states.

Food and Drug Administration. FDA regulations deal with materials which are additives to food. They constrain the kinds of chemistries of inks and coatings that may be used in direct contact with food. The primary impact is packaging and the guidelines are based on a positive list of materials. In the 1990s, concerns for materials that migrate to food is becoming an increasingly complex issue. The trend suggests that data on extractables in printed matter may become necessary to provide a basis for risk assessment of food contamination. This, in turn, will be an impetus to ink technologies that provide a

reduction in such extractables.

Resource Conservation and Recovery Act. The RCRA focuses on the proper disposition of waste from industrial processes. The interface to printing ink is primarily solvents, which can be flammable, and ingredients in ink that can contribute to the presence of certain heavy metals. The proper interface is the safe disposal of waste inks, but is often confused with disposal of printing matter.

Consumer Product Safety Commission. The CPSC establishes the maximum allowable levels of lead (qv) in children's toys and toy packaging. Outgrowths from these guidelines are the many voluntary toy manufacturers' guidelines limiting lead and other heavy metals. In general, heavy-metal specifications are an issue of the 1990s. Some guidelines limit total metals present while many constrain allowable soluble metal content by one of several standard protocols. These issues are emerging rapidly in Europe and Canada.

Coalition of Northeast Governors. The CONEG model heavy-metal guideline is implemented through state regulations and limits total metal content of lead, chromium, mercury, and cadmium. The limitation of 100 parts per million total is aimed at protecting the environment from the disposal of post-consumer waste.

Toxic Substances Control Act. The TSCA, a federal guideline, provides knowledge about the distribution of various chemicals into the society at large. It establishes a list known as the chemical inventory, and requires all printing ink products to be composed of raw materials that are on this TSCA inventory. A similar regulation is emerging in Canada and the European common market as well. Regulations such as this constrain new ink technology because of the testing requirements and the time frame required to list new chemicals.

Clean Air Act Amendments. The CAA define the components of an ink formula that are regarded as volatile organic compounds (VOCs). Emissions from ink formulations during the printing process are still considered to be a significant source of air pollution. This regulation continues to be the principal impetus to the introduction of ink technologies that offer alternatives to traditional solvent-based ink chemistries.

Clean Water Act. The CWA constrains all waste discharges to groundwater or waterways. Although not directly impacting printing ink making, these regulations pose an unexpected deterrent to some new ink technologies. The regulatory pressure to reduce volatile solvents emissions from flexographic and gravure printing processes has been the driving force to produce an extensive, advanced ink technology based on water-based chemistry. At the printing production level, water-based systems offer ease of clean-up of presses and fountains. Local sewage facilities have been used for disposal of such press wash and ink waste. New stringent effluent guidelines require pretreatment of press waste prior to discharge to sewage systems. This is a drawback to some printing segments seeking to use water-based technologies.

Economic Aspects

The world ink industry had total sales estimated to be about \$10.5 billion in 1992. The United States had about 32% (by weight) of world sales with the remaining portion split between Asia and Europe Africa. Japan has about 22% market share, Germany has 10%, and the U.K., Italy, and France each have 5%. The remaining share is split among many other countries.

The U.S. industry sold about \$2.92 billion in 1992. More than 60% of these sales were made by the top 10 producers. In order of size these were SunChemical, Flint Ink, INX, BASF, Zeneca, The Ink Co., Huber, Superior, Siegwerk, and Crown Zellerbach. These companies represent only 4% of the total number of ink companies in the United States.

BIBLIOGRAPHY

"Inks, Writing" in *ECT* 1st ed., Vol. 7, pp. 870–877, by R. S. Casey, W. A. Sheaffer Pen Co.; "Printing Ink" in *ECT* 1st ed., Vol. 11, pp. 149–163, by A. Voet, J. M. Huber Corp.; "Inks" in *ECT* 2nd ed., Vol. 11, pp. 611–632, by H. Dunn, B. V. Burachinsky, J. K. Ely, and P. W. Greubel, Interchemical Corp.; in *ECT* 3rd ed., Vol. 13, pp. 374–398, by B. V. Burachinsky, H. Dunn, and J. K. Ely, Inmont Corp.

1. S. P. Pappas, ed., *UV Curing: Science and Technology*, 2nd ed., Technology Marketing Corp., 1988.
2. E. A. Apps, ed., *Ink Technology for Printers and Students*, 3 Vols., Chemical Publishing Co., Inc., New York, 1964.
3. H. J. Wolfe, *Printing and Litho Inks*, 6th ed., MacNair-Dorland Co., New York, 1967.
4. J. Fetsko, ed., *Raw Material Data Handbooks*, Vols. 1–4, National Printing Ink Manufacturers Association, Hasbrouck Heights, N.J., 1983.
5. L. M. Larsen, *Industrial Printing Inks*, Reinhold Publishing Corp., New York, 1962.
6. C. A. Smith, *Mod. Paint Coat.* **69**(1), 30; (8), 47; (9), 61 (July, Aug., Sept. 1979).
7. T. C. Patton, ed., *Pigment Handbook*, 3 Vols., John Wiley & Sons, Inc., New York, 1979.
8. W. M. Morgans, *Pigments for Paints and Inks*, SITA Technology, London, 1977.
9. Wm. E. Rusterholz, *Ink Maker* **67**(11), 1z (1989).

General References

Printing Ink Handbook, National Association of Printing Ink Manufacturers, Hasbrouck Heights, N.J., 1988.

Rotogravure Ink, Gravure Technical Association and Champlain Co., Rochester, N.Y., 1957.

R. Leach, ed., *The Printing Ink Manual*, Van Nostrand Reinhold, Berkshire, U.K., 1988.

J. Fetsko, ed., *Relationship of Ink/Water Interactions to Printability of Lithographic Printing Inks*, part 1, National Printing Ink Manufacturers Association, Hasbrouck Heights, N.J., 1986.

J. Fetsko, ed., *Relationship of Ink/Water Interactions to Printability of Lithographic Printing Inks*, part 2, National Printing Ink Manufacturers Association, Hasbrouck Heights, N.J., 1988.

E. W. Zaleski, *TAGA Proceedings*, Rochester, N.Y., 1975, p. 345.

Ultraviolet/ Electron Beam Curing Formulations for Printing Ink, Coating and Paints, 3 Vols., SITA Technology, London, 1991.

D. Owen, *Printing Inks for Lithography*, 1990.

Wm. E. Rusterholz, *Focus on Inks*, High Volume Printing, Environment 2000, Apr. 1990.

R. W. Bassemir
A. Bean
O. Wasilewski
D. Kline
W. Hillis
C. Su
I. R. Steel
W. E. Rusterholz
Sun Chemical Corporation

INORGANIC HIGH POLYMERS

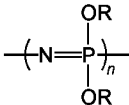
In a broad sense, inorganic polymers embrace a very large number of elements and compounds that are constructed of units bonded together in some type of repetitive fashion, with the requirement that the element carbon be excluded from the backbone of these repeating units. Some of the most common examples are the metals, as well as the naturally occurring silicate minerals and silica, where the SiO₄ tetrahedral moiety often forms the repeating unit of the one-, two-, or three-dimensional structure by sharing oxygen atoms with neighboring units (see SILICA). This article covers linear, mainly covalent macromolecules where the skeletal repeat unit is composed of primarily noncarbon elements.

The most commercially successful inorganic polymers to date are the polysiloxanes, owing to their unique high temperature stability, low temperature flexibility, and a number of other advantageous properties such as low surface energy and room-temperature vulcanizability (see SILICON COMPOUNDS, SILICONES). Despite the commercial success of polysiloxanes, however, the development of new inorganic polymers has not kept pace with that of their organic counterparts. The principal reasons for this are that a large variety of inorganic monomers has not been readily available, and extensive research and development with organic polymers have quickly provided many articles of commerce.

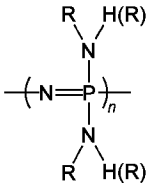
However, the rapid advancement of technology between the 1970s and the 1990s has created a need for materials whose special properties are unlikely to be obtained from carbon-based polymers alone. Thus, there has been a growing interest in the synthesis and development of new inorganic polymers, and there has been an almost explosive growth in the number of publications and patents dealing with the inorganic polymers, polyphosphazenes and polysilanes. These two polymer systems have been extensively studied, both with respect to new materials development via basic synthesis and chemical modification, and with respect to some of the intrinsic properties associated with the polymers; they are discussed here in detail. In addition, some newer inorganic polymers on the horizon are briefly covered. Other materials, eg, the polysilazanes, boron–nitrogen polymers, and other areas of inorganic polymers, are discussed elsewhere (1).

Polyphosphazenes

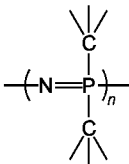
The polyphosphazenes, sometimes also referred to as polyphosphonitriles, are the most chemically versatile inorganic polymers known to date. A new class of phosphazene polymers, with only phosphorus–carbon bonded substituents, has been developed. Thus, based on their method of synthesis, two different types of phosphazene polymers are now in existence and are undergoing parallel development. The first type, bearing substituents on phosphorus bonded mostly via phosphorus–oxygen (1) and phosphorus–nitrogen (2) linkages, was developed in the mid-1960s as soluble, hydrolytically stable polymers (2–4). These polymers have seen steady growth in the diversity of polymers made throughout the 1970s and 1980s. Polymers of the second type, with substituents linked via direct phosphorus–carbon bonds (3) were first reported in the early 1980s (5–7) and since then have also seen significant development.



(1)



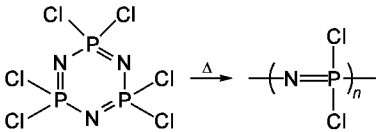
(2)



(3)

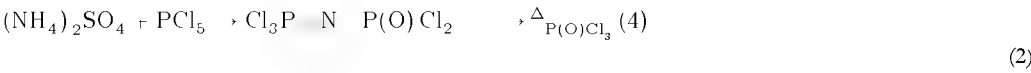
P≡O- AND P≡N-SUBSTITUTED POLYPHOSPHAZENES

Synthesis. The synthesis of poly(dichlorophosphazene) [25034-79-1], (N=PCL₂)_n (4), the parent polymer to over 300 macromolecules of types (1) and (2), is carried out via controlled, ring-opening polymerization of the corresponding cyclic trimer, (N=PCL₂)₃ [940-71-6].

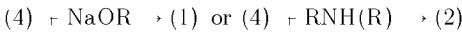


(4)

Various inorganic, organic, and organometallic compounds are known to catalyze this polymerization (4,8,9). Among these, BCl₃ is a very effective catalyst, although proprietary catalysts that significantly lower polymerization temperature from the usual, sealed-tube reaction at 250°C are involved in the industrial manufacture of the polymer. A polycondensation process has also been developed for the synthesis of (4) (10–12). This involves elimination of phosphoryl chloride from a monomer prepared from (NH₄)₂SO₄ and PCl₅.



The very high reactivity of the P–Cl bonds in (4) forms the basis for the now well-known macromolecular substitution method, which has been used to synthesize polymers of types (1) and (2) and some polymers that are hybrids of these and (3). The method involves nucleophilic reactions of (4), and to some extent of its difluoro analogue, with alkoxides or amines.



Properties. One of the characteristic properties of the polyphosphazene backbone is high chain flexibility which allows mobility of the chains even at quite low temperatures. Glass-transition temperatures down to –105°C are known with some alkoxy substituents. Symmetrically substituted alkoxy and aryloxy polymers often exhibit melting transitions if the substituents allow packing of the chains, but mixed-substituent polymers are amorphous. Thus the mixed substitution pattern is deliberately used for the synthesis of various phosphazene elastomers. On the other hand, as with many other flexible-chain polymers, glass-transition temperatures above 100°C can be obtained with bulky substituents on the phosphazene backbone.

The thermal stability of polymers of types (1) and (2) is also dependent on the nature of the substituents on phosphorus. Polymers with methoxy and ethoxy substituents undergo skeletal changes and degradation above about 100°C, but aryloxy and fluoroalkoxy substituents provide higher thermal stability (4). Most of the P–N- and P–O-substituted polymers either depolymerize via ring-chain equilibration or undergo cross-linking reactions at temperatures much above 150–175°C.

Phosphazene polymers are inherently good electrical insulators unless side-group structures allow ionic conduction in the presence of salts. This insulating property forms the basis for applications as wire and cable jackets and coatings. Polyphosphazenes also exhibit excellent visible and uv radiation transparency when chromophoric substituents are absent.

Another valuable characteristic of many phosphazene polymers is their flame-retardant behavior and low smoke generation on combustion (13). This property is utilized in commercial applications.

A remarkable feature of phosphazene polymers of types (1) and (2) is that appropriate substituents (which are readily attached) can be used as toggle switches to turn several properties, such as hydrolytic stability and electrical conductivity, on and off (1).

Applications. Among the P–O- and P–N-substituted polymers, the fluoroalkoxy- and aryloxy-substituted polymers have so far shown the greatest commercial promise (14–16). Both poly[bis(2,2,2-trifluoroethoxy)phosphazene] [27290-40-0] and poly(diphenoxyphosphazene) [28212-48-8] are microcrystalline, thermoplastic polymers. However, when the substituent symmetry is disrupted with a randomly placed second substituent of different length, the polymers become amorphous and serve as good elastomers. Following initial development of the fluorophosphazene elastomers by the Firestone Tire and Rubber Co., both the fluoroalkoxy (EYPEL-F) and aryloxy (EYPEL-A) elastomers were manufactured by the Ethyl Corp. in the United States from the mid-1980s until 1993 (see ELASTOMERS, SYNTHETIC PHOSPHAZENES).

The fluoroelastomers possess good rubber properties with the added advantages of being nonburning, hydrophobic, and solvent- and fuel-resistant. In addition to these, because of flexibility down to about –60°C, these polymers have been used in seals, gaskets, and hoses in army tanks, in aviation fuel lines and tanks, as well as in cold-climate oil pipeline applications. These polymers have also found application in various types of shock mounts for vibration dampening (14,17).

The aryloxyphosphazene polymers, on the other hand, have been used primarily in wire and cable coatings and jackets and as fire-resistant, low smoke, closed-cell foams and sound-barrier sheets.

Biomedical Applications. In the area of biomedical polymers and materials, two types of applications have been envisioned and explored. The first is the use of polyphosphazenes as bioinert materials for implantation in the body either as housing for medical devices or as structural materials for heart valves, artificial blood vessels, and catheters. A number of fluoroalkoxy-, aryloxy-, and arylamino-substituted polyphosphazenes have been tested by actual implantation in rats and found to generate little tissue response (18).

The second type of biomedical application utilizes the versatile chemistry of polyphosphazenes to generate bioactive polymers. Two approaches have been developed: one is to tie or physically entrap biologically active molecules using the phosphazene backbone as the carrier or encapsulant. The other is to attach bioactive molecules to a hydrolyzable (degradable) phosphazene backbone that releases the active species on breakdown of the backbone to harmless species that can be metabolized or directly excreted. Thus the first method has been used to attach a polymer-bound equivalent of the well-known anticancer agent cisplatin, heparin, dopamine, various enzymes, local anesthetics such as benzocaine, and a number of other bioactive molecules (19–26). The second approach utilizes cleverly designed polyphosphazenes that completely hydrolyze in water to small molecules. Thus, phosphazene polymers containing amino acid ester (eg, ethylglycinato) or imidazolyl substituents hydrolyze at body temperature and blood pH conditions to phosphate, ammonia, ethanol, and the corresponding amino acid or imidazole (27–30). The rate of hydrolysis can often be controlled by the presence of another substituent on phosphorus that does not allow ready hydrolysis. In this manner, bioactive agents can be released in a controlled fashion. Successful release of steroids, the antitumor agent melphan, and of naproxen has been obtained in *in vitro* and *in vivo* studies (31–36). The attachment of oligopeptides to a specially designed side-group on polyphosphazenes has also been reported (37), and could lead to the development of useful biomaterials.

Two crucial aspects of the design of bioactive polyphosphazenes have been carefully developed. One involves the hydrophilicity or hydrophobicity of the polymer, and the other is the stability of the polymer or tactical substituent linkages that allow release of the active agent or ensure its potency to be retained in the bound form. For example, methylamino-, glyceryl-, glucosyl-, (hydroxyalkyl)amino-, and alkyl ether-substituted polyphosphazenes are water-soluble, whereas fluoroalkoxy or aryloxy phosphazenes are hydrophobic. On the other hand, chemical bonds like Schiff's base linkages are hydrolytically unstable. Thus, the release rate or activity of bioactive molecules attached to a phosphazene polymer can be fine-tuned to obtain desired effects by balancing substituent characteristics on the polymer and the nature of the linkage used to attach the bioactive species. The general structure in Figure 1a shows the chemical architectural characteristics that have been employed in the design of these polymers and 1b exemplifies some bioactive molecules that have been attached to the phosphazene backbone through an amido linkage, in each case.

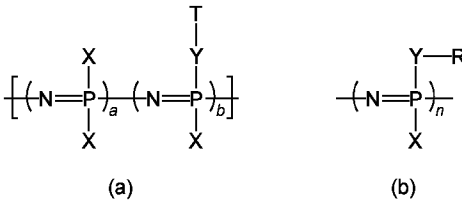
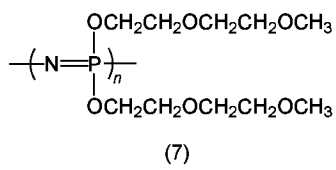
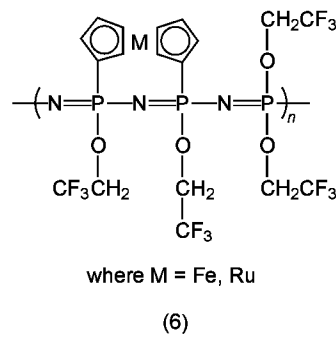


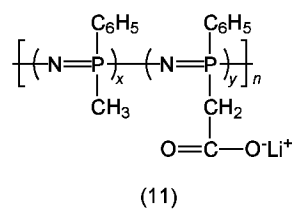
Fig. 1. Schematic bioactive polyphosphazenes. (a) General structure, where X = hydrophilic–hydrophobic group that hydrolyzes with concurrent polymer breakdown, Y = difunctional group for attaching bioactive agent to polymer, and T = bioactive agent. (b) Actual example where X = –OC₂H₅, Y = and R = (CH₃)₂C(SH)–CH(NHCOCH₃), 4-(C₃H₇)₂NSO₂–C₂H₅, etc.

Polymers Bearing Metal Complexes. A large number of polymers with side groups containing metal complexes have been reported. The

complexes are linked to the phosphazene backbone primarily through a ligand on a substituent (38–42), although linkages through the skeletal nitrogen (19) or through direct metal–phosphorus bonding with the skeletal phosphorus atoms (43) have also been utilized. The impetus for the synthesis of these metal–polymer hybrids is the potential for generating materials using particular properties of metals such as catalysis, electrical conduction, magnetism, and in some instances biological activity, with the inorganic polymer support providing features such as solubility, chemical protection, control of activity, biocompatibility, prevention of leaching, etc. Biological activity is exhibited by the platinum complex cisplatin, and catalytic activity has been demonstrated with polymer-bound species such as (5). Polymers (5) and (6) are representative of metal-containing polyphosphazenes where the metal is linked through a ligand on a side-group.



Solid Electrolyte Applications. Among other potentially useful polymers synthesized by the versatile macromolecular substitution process are polymers based on oligoether substituents or heterocyclic substituents that have been under intense investigation for solid electrolyte battery applications (44–47). The most promising of these is poly[bis(methoxyethoxyethoxy)phosphazene] (MEEP) (7). This polymer is an excellent solid solvent for salts like lithium or silver triflate, and the resulting solid solution has an ionic conductivity three orders of magnitude greater than a similar composition based on poly(ethylene oxide). Lightweight, prototype rechargeable batteries based on cross-linked (7) have been successfully tested and are approaching commercialization.



A number of liquid crystalline polyphosphazenes with mesogenic side groups have been prepared (48–50). Polymers with nonlinear optical activity have also been reported (51). Polyphosphazene membranes have been examined for gas, liquid, and metal ion separation, and for filtration (52–54). There is interest in phosphazene–organic copolymers, blends, and interpenetrating polymer networks (IPNs) (55–61) to take advantage of some of the special characteristics of phosphazenes such as flame retardance and low temperature flexibility. A large number of organic polymers with cyclophosphazene substituents have been made (62).

POLYMERS WITH ALKYL AND ARYL SUBSTITUENTS ON P

Even though partially alkyl- and aryl-substituted polyphosphazenes are accessible via the ring-opening polymerization followed by the macromolecular substitution route (63), polymers in which all substituents are attached through direct phosphorus–carbon bonds (3) are not yet accessible by this method. Reaction of organolithium or organomagnesium compounds with polymers such as polydichlorophosphazene (4) leads to incompletely substituted polymer as well as polymer degradation, and fully alkyl aryl-substituted cyclic phosphazenes do not undergo ring-opening polymerization (4).

Synthesis. The first fully alkyl aryl-substituted polymers were reported in 1980 via a condensation–polymerization route. The method involves, first, the synthesis of organophosphine-containing alkyl or aryl substituents, followed by the ready oxidation of the phosphine to a phosphorane with leaving groups suitable for a 1,2-elimination reaction. This phosphorane is then thermally condensed to polymers in which all phosphorus atoms bear alkyl or aryl substituents. This condensation synthesis is depicted in Figure 2 (5–7,64).

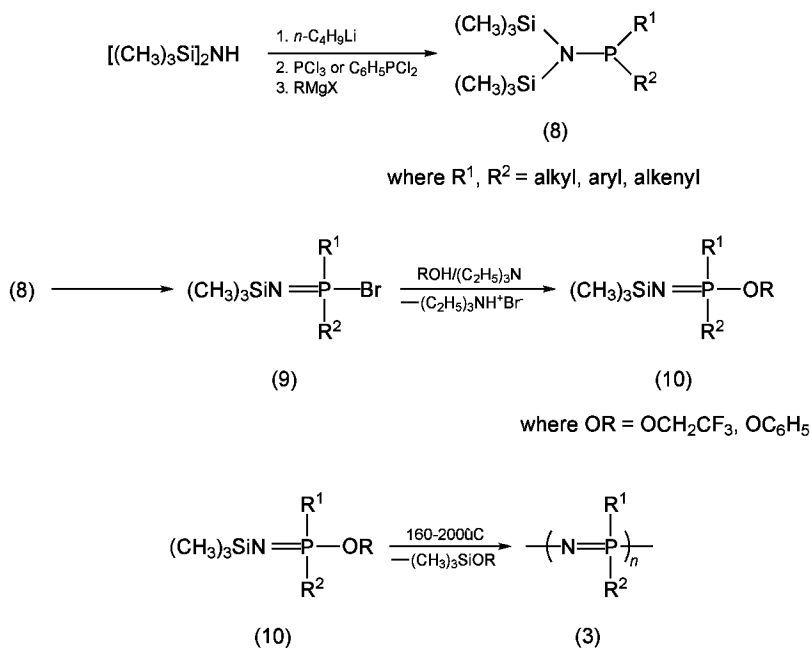
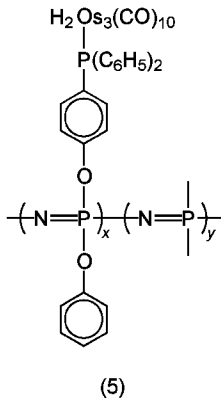


Fig. 2. Synthesis of polyphosphazenes with P—C bonds.

The three steps in equation 3 are carried out in one vessel. This affords a wide variety of disilylaminoorganophosphines (8), including those with vinyl substituents (65), in yields of 40–85%. The oxidation of (8) to (9) and the reaction of (9) with alcohol (eq. 4) are carried out in a second reactor to provide the "monomer" phosphoranimines (10) in overall 30–65% yield based on starting PCl₃ or C₆H₅PCl₂. The use of C₂Cl₄ in place of Br₂ in the conversion of (8) to (9) makes it possible to carry out all the reactions leading to (10) in one vessel, and this has significantly increased yields of the monomer, with overall yields up to 80% (66).

The bulk polycondensation of (10) is normally carried out in evacuated, sealed vessels such as glass ampules or stainless steel Parr reactors, at temperatures between 160 and 220°C for 2–12 d (67). Two monomers with different substituents on each can be cocondensed to yield random copolymers. The by-product silyl ether is readily removed under reduced pressure, and the polymer purified by precipitation from appropriate solvents. Catalysis of the polycondensation of (10) by phenoxide ion in particular, as well as by other species, has been reported to bring about complete polymerization in 24–48 h at 150°C (68). Catalysis of the polycondensation of phosphoranimines that are similar to (10), but which yield P—O-substituted polymers (1), has also been described and appears promising for the synthesis of (1) with controlled structures (69,70).

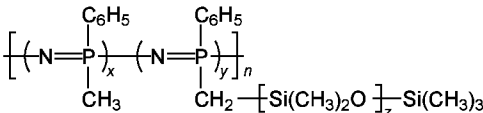
In addition to providing fully alkyl-aryl-substituted polyphosphazenes, the versatility of the process in Figure 2 has allowed the preparation of various functionalized polymers and copolymers. Thus the monomer (10) can be derivatized via deprotonation–substitution, when a *P*-methyl (or P—CH₂–) group is present, to provide new phosphoranimines some of which, in turn, serve as precursors to new polymers (64). In the same vein, polymers containing a P—CH₃ group, for example, poly(methylphenylphosphazene), can also be derivatized by deprotonation–substitution reactions without chain scission. This has produced a number of functionalized polymers (64,71–73), including water-soluble carboxylate salts (11), as well as graft copolymers with styrene (74) and with dimethylsiloxane (12) (75).



REPLACEThus alkyl- and aryl-substituted polyphosphazenes and their immediate precursors are also quite amenable to synthetic modifications, with the potential for the synthesis of a wide variety of materials being quite evident.

Properties. The condensation–polymerization reaction (eq. 5) yields alkyl- and aryl-substituted polymers with average molecular weights in the range 40,000 to 250,000 (*M_n* ranges from 20,000 to 100,000). In addition to routine size exclusion chromatography (sec), membrane osmometric, viscometric, and light-scattering experiments have been used in representative cases to determine absolute molecular weights and to obtain polymer chain dimensions (76). In general, the polymers are soluble in chlorinated solvents such as CH₂Cl₂ and CHCl₃. Polymers with phenyl substituents are also soluble in tetrahydrofuran.

The P—N backbone remains quite flexible with small, unbranched alkyl substituents on phosphorus. For example, polydimethylphosphazene (13) exhibits a glass-transition temperature (*T_g*) at –46°C. Bulkier substituents like phenyl cause significant increases in *T_g*, however. Symmetrically substituted dialkyl polymers exhibit varying degrees of crystallinity by x-ray diffraction (76). Polymer (13) shows both amorphous and crystalline domains, whereas the diethyl analogue (14) is so highly crystalline that it is insoluble in all common solvents, and does not exhibit a glass-transition temperature. Asymmetrically substituted polymers such as poly(methylphenylphosphazene) (15) are amorphous, which is probably a result of their atactic structure (76).REPLACE



(12)

Alkyl- and aryl-substituted polyphosphazenes (3) exhibit onset of decomposition at between 350 and 400°C, as determined by thermogravimetric analyses (tga) (76) under argon atmosphere. Interestingly, the decomposition temperature for the graft copolymer (12) (Table 1) is higher than that of either of the component homopolymers. Even though it had long been predicted that the P—C bonded polymers (3) were likely to be more thermally stable than the P—O- (1) and P—N-bonded (2) polymers, a detailed study (77) suggests there may not be significant differences between the thermal stabilities of P—O- and P—C-bonded polymers. However, information on the thermooxidative stability of (3) as cross-linked polymers, in filled or unfilled systems, is not yet available.

Table 1. Thermal Analytical Data^a for Some Poly(alkyl/arylphosphazenes)

Polymer	Structure number	CAS Registry Number	<i>T_g</i> , °C	<i>T_g</i> , °C	<i>T_{onset}</i> , °C
polydimethyl-phosphazene	(13)	[88718-77-8]	46	143	401
polydiethyl-phosphazene	(14)	[107037-73-0]		217	353
poly(methylphenyl-phosphazene)	(15)	[88718-66-5]	37		354
poly(dimethylphos-phazene- <i>co</i> -methyl-phenylphosphazene)	(16)	[96743-50-9]	3		390
polymer salt			51 ^b		320
poly(phosphazene- <i>g</i> -siloxane)	(12)		123, 38	45 ^c	436 ^d

^a *T_g* and *T_m* by dsc, *T_{onset}* (onset of decomposition) Thermogravimetric analysis in argon at 10°C heating rate.

^b Estimate.

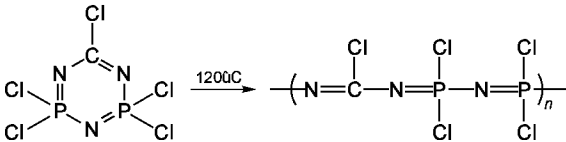
^c Melting transition for poly(dimethylsiloxane) graft segment.

^d In air.

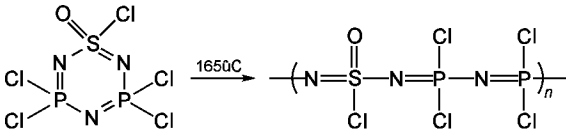
Applications. Polymers with small alkyl substituents, particularly (13), are ideal candidates for elastomer formulation because of quite low temperature flexibility, hydrolytic and chemical stability, and high temperature stability. The ability to readily incorporate other substituents (in addition to methyl), particularly vinyl groups, should provide for conventional cure sites. In light of the biocompatibility of polysiloxanes and P—O- and P—N-substituted polyphosphazenes, poly(alkyl arylphosphazenes) are also likely to be biocompatible polymers. Therefore, biomedical applications can also be envisaged for (3). A third potential application is in the area of solid-state batteries. The first steps toward ionic conductivity have been observed with polymers (13) and (15) using lithium and silver salts (78).

PHOSPHAZENES CONTAINING SKELETAL CARBON, SULFUR, AND METAL ATOMS

The first phosphazene polymers containing carbon (79), sulfur (80,81), and even metal atoms (82) in the backbone have been reported. These were all prepared by the ring-opening polymerization of partially or fully chloro-substituted (or fluoro-substituted) trimers containing one hetero atom substituting for a ring-phosphorus atom in a cyclotriphosphazene-type ring.

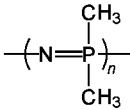


(16)

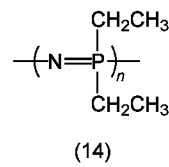


(17)

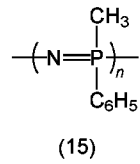
Apparent ring strain in these trimers of a magnitude greater than in the corresponding cyclotriphosphazene causes the former to polymerize at significantly lower temperatures. From the halo-substituted polymers, aryloxy-substituted polymers have been made for the carbon- and sulfur-containing systems, and these polymers show glass-transition temperatures (*T_g*) higher than the corresponding phosphazene homopolymers when carbon is a skeletal atom, but lower or higher *T_g* (depending on the nature of the aryloxy group) when sulfur(VI) is a skeletal atom. The copolymers containing skeletal sulfur(VI) substitution appear especially promising from a materials standpoint. Equations 6 and 7 are representative of the ring-opening polymerization route used to prepare the heterophosphazene polymers. Macromolecular substitution, eg, on (16) and (17), then yields (18) and (19), respectively.



(13)

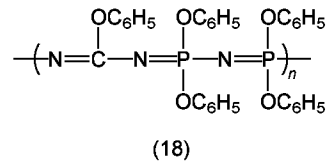


An example of a polyphosphazene incorporating metal atoms is (20), where M = Mo or W.

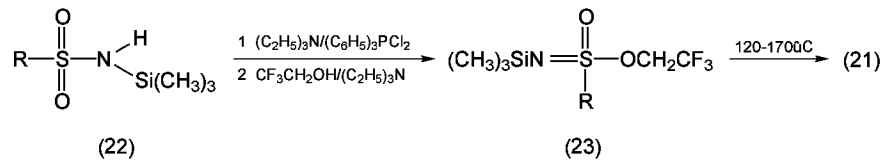


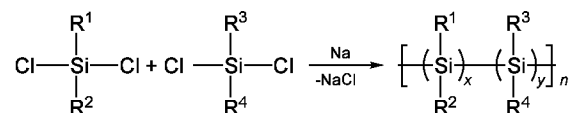
Poly(alkyl/aryloxothiazenes)

The synthesis of a new class of inorganic polymers (21) with a backbone consisting of alternating sulfur(VI) and nitrogen atoms, and with variable alkyl or aryl substituents as well as a fixed oxygen substituent on sulfur, has recently been accomplished (83–85). These polymers are structurally analogous to poly(alkyl arylphosphazenes).



Synthesis and Properties. The synthesis of (21) follows a very straightforward route based on readily accessible starting materials and on some novel reactions in organo–inorganic sulfur chemistry (83–85), as well as on polycondensation chemistry analogous to that utilized in the preparation of poly(alkyl arylphosphazenes). One preparation of (21) is as follows:





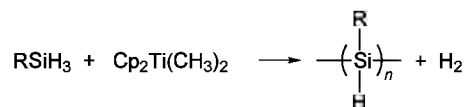
One of the constraints of this method is that only a limited number of functional substituents can be introduced because of the high reactivity of the alkali metal used in the coupling process. Despite this, a large number of homo- and copolymers have been made just from alkyl aryl-substituted dichlorosilanes. Some polymers and copolymers containing heteroatoms, as well as fluorinated polymers, have also been reported (94–97). The sodium-coupling process is an extremely complex heterogeneous reaction, with profound effects on such properties as polymer yield, molecular weight, polydispersity, and the like brought about by the order of metal and chlorosilane addition, the nature of the alkali metal, the polar or nonpolar nature of the solvent used, the reaction temperature, polar additives used, and a number of other factors. Studies of these issues have been reviewed (94). An earlier article, though much smaller in scope, provides a very precise and comprehensible summary of the chemistry, properties, and potential applications of polysilanes (98).

In general, the sodium coupling reaction (often carried out in a refluxing hydrocarbon, such as toluene) produces a mixture of polymer, oligomers, and cyclic species, with polymer yields in the low to moderate range. The polymer molecular weight distribution is often bimodal or trimodal and broad, but this process produces high molecular weight material (average M_w can be as high as several million) that is not yet accessible via other routes. Polar solvents such as tetrahydrofuran as well as diethyl ether (99) and additives such as diglyme, crown ethers (100), and, more recently, ethyl acetate (101) have been shown to provide increased yields of polymers with lower polydispersity and unimodal distribution, although depending on specific conditions molecular weights can be significantly lower. Ultrasound has also been used to improve polymer yields (94,102).

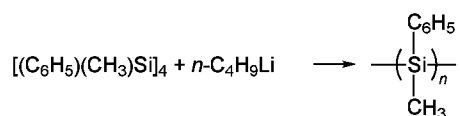
The mechanism of the polycondensation reaction remains unclear. A variety of possible reactive intermediates have been suggested, including silyl radicals and silyl anions. An anionic propagation mechanism (100,101,103) has been strongly suggested, although the case is by no means settled (104).

Other Synthetic Methods.

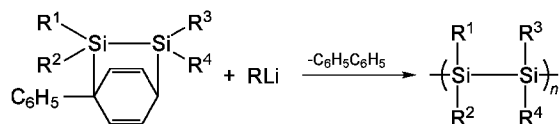
Dehydrogenative Coupling. Transition-metal catalyzed polymerization of silanes appears to hold promise as a viable route to polysilanes. A number of transition-metal complexes have been investigated, with titanium and zirconium complexes being the most promising (105–108). Only primary silanes are active toward polymerization, and molecular weights are rather low. The dehydrogenative polymerization is depicted in reaction 11, where Cp = cyclopentadienyl:



Ring-Opening Polymerization. As with most other inorganic polymers, ring-opening polymerization of cyclotetrasilanes has been used to make polysilanes (109,110). This method, however, has so far only been used for polymethylphenylsilane (eq. 12). Molecular weights (up to 100,000) are higher than from transition-metal catalyzed polymerization of primary silanes.



Polymerization of Masked Disilenes. A novel approach, namely, the anionic polymerization of masked disilenes, has been used to synthesize a number of poly(dialkylsilanes) as well as the first dialkylamino substituted polysilanes (eq. 13) (111,112). The route is capable of providing monodisperse polymers with relatively high molecular weight ($M_n = 10^4 - 10^5$), and holds promise of being a good method for the synthesis of alternating and block copolymers.



Electrochemical Synthesis. Electrochemical methods have also been investigated for the synthesis of polysilanes, but these have so far yielded low molecular weight materials (113,114).

Polymer Modification. The introduction of functional groups on polysilanes using the alkali metal coupling of dichlorosilanes is extremely difficult to achieve. Some polymers and copolymers with 2-(3-cyclohexenyl)ethyl substituents on silicon have been made, and these undergo hydrogen halide addition to the carbon–carbon double bond (94,98).

An easier approach to functionalizing polysilanes utilizes the ready cleavage of phenyl groups from silicon in polymethylphenylsilane using a strong acid such as HCl or, better, triflic acid (98,115,116). The resulting polymers containing Si–Cl or Si–triflate linkages can then be used as electrophilic substrates for attaching various functional groups to silicon.

The functionalization of poly(phenylsilane) [99936-07-9] by reaction with CCl_4 and with CBr_4 has also been reported (117). This yields polymers containing Si–Cl or Si–Br bonds, but leaves the Si–C–H bonds intact.

PROPERTIES

As with many polymer systems, the properties of polysilanes vary considerably depending on the nature of the substituents on the silicon atoms in the backbone. Polymers symmetrically substituted with the smallest alkyl groups are highly crystalline and, therefore, insoluble. This is also true of poly(diarylsilanes), unless crystallinity is disrupted by the introduction of appropriate substituents on the ring (118,119). However, most unsymmetrically substituted dialkyl and alkyl aryl homopolymers as well as copolymers are soluble in solvents such as tetrahydrofuran or toluene. The longer Si–Si bond length, compared with the C–C bond length, allows quite a bit of flexibility in the backbone such that glass-transition temperatures as low as 76°C (for poly(*n*-hexylmethylsilane)), have been observed. On the other hand, as expected, aryl substitution brings about significant increases in T_g . Thus polysilanes cover the range from rubbery elastomers to brittle solids. Thermal stability studies on polysilanes have not been carried out in any systematic fashion, but most polymers start to degrade in the temperature range 250–300°C (94).

Polysilanes are chemically inert to air and water at ordinary temperature, but their reactivity increases in solvent. In a solvent such as tetrahydrofuran, degradation of the Si–Si backbone by strong bases is quite rapid. Strong oxidizing agents like *m*-chloroperbenzoic acid insert oxygen atoms between the silicons to produce Si–O–Si linkages in the backbone (120).

Electronic Properties. What distinguishes polysilanes from virtually all other polymers is their backbone σ -conjugation. This leads to strong electronic absorption in the near-uv from a σ – σ^* transition. For most homo- and copolymers the absorption maximum (λ_{max}) lies between 300 and 400 nm. Dialkyl substituted polymers absorb in the 300–325 nm range in solution (94). A fluoroalkyl polymer with a shorter wavelength absorption at 285 nm has been reported (96). Aryl alkyl and diaryl substituted homo- and copolymers, on the other hand, absorb between 330 and 400 nm. The lowering of

transition energy with aryl substituents has been attributed to the mixing of aryl- π and SiSi- σ orbitals (94,98). Both the absorption maximum and the extinction coefficient (per Si—Si bond) increase till the chain length reaches about 40–50 silicon atoms (94). Some representative uv absorption data are provided in Table 2. Many polysilane homopolymers exhibit thermochromism (as well as piezochromism) in their electronic absorption behavior (94,98,121,122), both in solution and in the solid state (see CHROMOGENIC MATERIALS). Significant red shifts are often observed in variable temperature uv spectra of the polymers. This lowering of transition energy has been attributed primarily to increases in the trans arrangement of the Si—Si bonds in the backbone at lower temperature, although other structural factors are also thought to be involved. Understanding of the dependence of electronic transition in polysilanes on conformation of the backbone as well as on intra- and intermolecular forces is still evolving (123).

Table 2. Ultraviolet Absorption of Polysilanes in Solution

Polymer	CAS Registry Number	λ_{max} , nm
$(n\text{-C}_3\text{H}_7\text{SiCH}_3)_n$	[109926-38-7]	306
$(n\text{-C H}_1\text{ }_3\text{SiCH}_3)_n$	[88002-83-9]	306
(cyclohexyl SiCH ₃) _n	[88002-85-1]	326
$(n\text{-C}_{12}\text{H}_{25}\text{SiCH}_3)_n$	[88002-84-0]	309
$(\text{CF}_3\text{CH}_2\text{CH}_2\text{SiCH}_3)_n$	[31324-76-2]	285
$[(n\text{-C}_4\text{H}_9)_2\text{Si}]_n$	[97036-65-2]	314
$[(n\text{-C H}_1\text{ }_3)_2\text{Si}]_n$	[97036-67-4]	317
$[(n\text{-C}_{10}\text{H}_{21})_2\text{Si}]_n$	[117652-57-0]	324
$(\text{C H}_5\text{SiCH}_3)_n$	[76188-55-1]	335
$(\text{CH}_3\text{O—C H}_4\text{—SiCH}_3)_n$	[97464-14-7]	337
$[(\text{CH}_3)_2\text{N—C H}_4\text{—SiCH}_3]_n$	[122644-41-1]	361
$[(n\text{-C}_4\text{H}_9\text{—C H}_4\text{—})_2\text{Si}]_n$	[111939-58-3]	395
$\{[(n\text{-C H}_1\text{ }_3)_2\text{Si}]_{1\text{ }17}[(\text{CH}_3)_2\text{Si}]_1\}_n$	[118338-24-2]	310
$\{[(\text{CH}_3)_2\text{Si}]_{1\text{ }13}[(\text{C H}_5)_2\text{Si}]_1\}_n$	[70926-75-9]	351

The polysilanes are normally electrical insulators, but on doping with AsF₅ or SbF₅ they exhibit electrical conductivity up to the levels of good semiconductors (qv) (98,124). Conductivities up to 0.5 (Ω·cm)^{–1} have been measured. However, the doped polymers are sensitive to air and moisture thereby making them unattractive for practical use. In addition to semiconducting behavior, polysilanes exhibit photoconductivity and appear suitable for electrophotography (qv) (125–127). Polysilanes have also been found to exhibit nonlinear optical properties (94,128). Polysilanes absorb electromagnetic energy and undergo chain scission (94,98,129). This is an extremely important property of these polymers in terms of applications. Photochemistry is exhibited both in solution and in the solid state. The quantum yields for photocission in solution are quite high, ranging from 0.5 to 1.0. Some cross-linking is observed with aryl-substituted polymers, but even with these scission predominates. During irradiation (for example, with 254-nm photons), the absorption maximum for a polymer is gradually blue shifted and decreases in intensity due to breakdown of the polymer into fragments with increasingly smaller catenation, with a corresponding decrease in backbone σ -conjugation. This has been termed photobleaching and is a key characteristic of polysilanes for microlithographic applications. Photolysis also occurs with uv photons of wavelength over 300 nm.

APPLICATIONS

Manufacture of β -Silicon Carbide. A commercially utilized application of polysilanes is the conversion of some homopolymers and copolymers to silicon carbide (130). For example, polydimethylsilane is converted to the ceramic in a series of thermal processing steps. Silicon carbide fibers is commercialized by the Nippon Carbon Co. under the trade name Nicalon (see REFRACTORY FIBERS). **Microlithography, Xerography.** Because of their photosensitivity, polysilanes are under intense investigation for use as positive photoresist materials (94) (see LITHOGRAPHIC RESISTS). They are particularly attractive because both wet and dry development techniques can be used for imaging (131,132). The use of polysilanes for xeroprinting has been reported (133). Thermal and optical sensors based on the photodegradation of polysilanes have been developed (134). **Photoinitiation.** Since photolysis of polysilanes generates silyl radicals, which can add to carbon—carbon double bonds, these polymers have been used for the free-radical polymerization of unsaturated organic monomers (135,136). Though about one-tenth as efficient as other organic photoinitiators, polysilanes are nevertheless quite insensitive to oxygen effects, which somewhat compensates for their lower efficiency.

Polygermanes

Soluble and well-characterized polygermane homopolymers, (R₂Ge)_n, and their copolymers with polysilanes have been prepared by the alkali metal coupling of diorgano-substituted dihalogermanes (137–139), via electrochemical methods (140), and by transition-metal catalyzed routes (105), as with the synthesis of polysilanes. The polygermanes exhibit many of the same electronic properties as polysilanes, including near-uv photoabsorption, thermochromism, photobleaching, as well as nonlinear optical activity, and have seen a fair amount of theoretical and experimental investigation (137,138,141–143). However, despite similarities with polysilanes, polygermanes appear to be unlikely candidates for commercial exploitation, mainly because of higher monomer costs for germanium compounds and the lack of clear superiority over polysilanes in any area of application sufficient to offset the cost disadvantage.

BIBLIOGRAPHY

"Inorganic High Polymers" in *ECT* 2nd ed., Vol. 11, pp. 632–650, by B. P. Block, Pennsalt Chemicals Corp.; in *ECT* 3rd ed., Vol. 13, pp. 398–413, by E. N. Peters, Union Carbide Corp.

1. J. E. Mark, H. R. Allcock, and R. West, *Inorganic Polymers*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1992.
2. H. R. Allcock and R. L. Kugel, *J. Am. Chem. Soc.* **87**, 4216 (1965).
3. H. R. Allcock, R. L. Kugel, and K. J. Valan, *Inorg. Chem.* **5**, 1709 (1966).
4. H. R. Allcock, *Phosphorus–Nitrogen Compounds*, Academic Press, Inc., New York, 1972.
5. P. Wisian-Neilson and R. H. Neilson, *J. Am. Chem. Soc.* **102**, 2848 (1980).
6. R. H. Neilson and P. Wisian-Neilson, *J. Macromol. Sci. Chem.* **A16**, 425 (1981).
7. **U.S. Pat. 4,412,053** (1983), R. H. Neilson and P. J. Wisian-Neilson.
8. G. L. Hagnauer, *J. Macromol. Sci. Chem.* **A16**, 385 (1981).
9. H. R. Allcock, in *Catalytic Polymer Synthesis*, ACS Symposium Series, Vol. 496, American Chemical Society, Washington, D.C., 1992, p. 236.

10. **Fr. Pat. 466 435** (1981), R. De Jaeger, M. Heloui, and E. Puskaric.
11. Ph. Potin and R. De Jaeger, *Eur. Polym. J.* **27**, 341 (1991), and references therein.
12. G. D'Halluin, R. De Jaeger, J. P. Chambrette, and Ph. Potin, *Macromolecules* **25**, 1254 (1992).
13. P. J. Lieu, J. H. Magill, and Y. C. Alarie, *J. Fire Flammabil.* **11**, 167 (1980).
14. H. R. Penton, in M. Zeldin, H. R. Allcock, and K. J. Wynne, eds., *Inorganic and Organometallic Polymers*, ACS Symposium Series, Vol. 360, American Chemical Society, Washington, D.C., 1988, p. 277.
15. S. H. Rose, *J. Polym. Sci.* **B6**, 837 (1968).
16. D. F. Lohr and J. A. Beckman, *Rubber Plast. News* **16**, (1982).
17. Product lit. no. SC-20A (987), Ethyl Corp., Baton Rouge, La., 1987.
18. C. W. R. Wade and co-workers, in C. E. Carraher, J. E. Sheats, and C. U. Pittman, eds., *Organometallic Polymers*, Academic Press, Inc., New York, 1978, pp. 289–300.
19. H. R. Allcock, R. W. Allen, and J. P. O'Brien, *J. Am. Chem. Soc.* **99**, 3984 (1977).
20. T. X. Neenan and H. R. Allcock, *Biomaterials* **3**, 2 (1982).
21. *Ibid.*, p. 78.
22. H. R. Allcock, W. C. Hymer, and P. E. Austin, *Macromolecules* **16**, 1401 (1983).
23. H. R. Allcock and S. Kwon, *Macromolecules* **19**, 1502 (1986).
24. H. R. Allcock, P. E. Austin, and T. X. Neenan, *Macromolecules* **15**, 689 (1982).
25. S. Cohen and co-workers, *J. Am. Chem. Soc.* **112**, 7832 (1990).
26. H. R. Allcock, in *Inorganic Polymers*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1992, pp. 95–118.
27. H. R. Allcock, T. J. Fuller, D. P. Mack, K. Matsumura, and K. M. Smeltz, *Macromolecules* **10**, 824 (1977).
28. H. R. Allcock, T. J. Fuller, and K. Matsumura, *Inorg. Chem.* **21**, 515 (1982).
29. T. J. Fuller and H. R. Allcock, *J. Am. Chem. Soc.* **103**, 2250 (1981).
30. J. H. L. Krommen, J. Vandorpe, and E. H. Schacht, *Macromol. Chem.* **193**, 1261 (1992).
31. H. R. Allcock and T. J. Fuller, *Macromolecules* **13**, 1338 (1980).
32. C. Laurencin and co-workers, *J. Biomed. Mater. Res.* **21**, 1231 (1987).
33. C. W. J. Grolleman and co-workers, *J. Controlled Release* **3**, 143 (1986); **4**, 119 (1986); **4**, 133 (1986).
34. J. H. Goedemoed and K. De Groot, *Makromol. Chem. Macromol. Symp.* **19**, 341 (1988).
35. **World Pat. 9,205,778** (1992), S. Cohen and co-workers.
36. J. H. Goedemoed and co-workers, *J. Controlled Release* **17**, 235 (1991).
37. H. R. Allcock and J. Y. Chang, *Macromolecules* **24**, 993 (1991).
38. H. R. Allcock, T. J. Fuller, and T. L. Evans, *Macromolecules* **13**, 1325 (1980).
39. H. R. Allcock, K. D. Lavin, N. M. Tollefson, and T. L. Evans, *Organometallics* **2**, 267 (1983).
40. H. R. Allcock, G. H. Riding, and K. D. Lavin, *Macromolecules* **20**, 6 (1987).
41. H. R. Allcock, I. Manners, M. N. Mang, and M. Parvez, *Inorg. Chem.* **29**, 522 (1990).
42. H. R. Allcock, J. L. Desorcie, and G. H. Riding, *Polyhedron* **6**, 119 (1987).
43. H. R. Allcock, M. N. Mang, G. S. McDonnell, and M. Parvez, *Macromolecules* **20**, 2060 (1987).
44. J. L. Bennett, A. A. Dembek, H. R. Allcock, B. J. Heyen, and D. F. Shriver, *Chem. Mater.* **1**, 14 (1989).
45. G. Nazri, D. M. MacArthur, and J. F. O'Gara, *Chem. Mater.* **1**, 370 (1989).
46. J. M. G. Cowie, *Makromol. Chem. Macromol. Symp. (Int. Symp. Polym. Mater.—Prep. Charact. Prop. 1991)* **53**, 43 (1992).
47. H. R. Allcock, J. A. Dodge, L. S. Van Dyke, and C. R. Martin, *Chem. Mater.* **4**, 780 (1992).
48. R. E. Singler, R. A. Willingham, R. W. Lenz, A. Furakawa, and H. Finkelmann, *Macromolecules* **20**, 1727 (1987).
49. R. E. Singler and co-workers, in R. A. Weiss and C. K. Ober, eds., *Liquid Crystalline Polymers*, ACS Symposium Series, American Chemical Society, Vol. 435, Washington, D.C., 1990, p. 185.
50. H. R. Allcock and C. Kim, *Macromolecules* **23**, 3881 (1990).
51. H. R. Allcock and co-workers, *Macromolecules* **24**, 1000 (1991).
52. M. L. Stone and co-workers, in M. K. Turner, ed., *Eff. Ind. Membr. Processes: Benefits Oppor.*, Elsevier, London, 1991, p. 321.
53. **U.S. Pat. 5,066,398** (Nov. 19, 1991), R. Soria, C. Defalque, and J. Gillot.
54. E. Drioli, G. Golemme, A. Basile, and S. M. Zhang, *Key Eng. Mater. (Inorg. Membr. ICIM₂-91)*, 611 (1991).
55. **Jpn. Pat. 03/296546** (Dec. 27, 1991), T. Takano.
56. **Ger. Pat. 4,026,883** (Feb. 27, 1992), H. Engelhard, G. Spielgies, and G. Tuerck.
57. M. Gleria and co-workers, *Chem. Mater.* **4**, 1027 (1992).
58. B. K. Coltrain and co-workers, *Chem. Mater.* **4**, 358 (1992).
59. H. R. Allcock and K. B. Visscher, *Chem. Mater.* **4**, 1182 (1992).
60. H. R. Allcock, K. B. Visscher, and I. Manners, *Chem. Mater.* **4**, 1188 (1992).
61. A. V. Semakov and co-workers, *Vysokomol. Soedin. Ser. B* **34**, 29 (1992).
62. C. W. Allen, *Stud. Inorg. Chem.* (Chem. Inorg. Ring. Syst.) **14**, 171 (1992).
63. H. R. Allcock, *Stud. Inorg. Chem.* (Chem. Inorg. Ring. Syst.) **14**, 145 (1992).
64. R. H. Neilson and P. Wisian-Neilson, *Chem. Rev.* **88**, 541 (1988).
65. K. E. Waters and R. H. Neilson, *Phos. Sulf.* **39**, 189 (1988).
66. R. H. Neilson, in press (1994).
67. P. Wisian-Neilson and R. H. Neilson, *Inorg. Synth.* **25**, 69 (1988).
68. R. H. Neilson and co-workers, *Polym. Preprints* **34**(1), 263 (1993).
69. K. Matyjaszewski, *J. Inorg. Organomet. Polym.* **2**, 5 (1992).
70. K. Matyjaszewski, M. S. Lindberg, M. K. Moore, and M. L. White, *Polym. Preprints* **34**(1), 274 (1993).
71. P. Wisian-Neilson, R. R. Ford, R. H. Neilson, and A. K. Roy, *Macromolecules* **19**, 2089 (1986).
72. P. Wisian-Neilson and R. R. Ford, *Macromolecules* **22**, 72 (1989).
73. P. Wisian-Neilson and co-workers, *Macromolecules* **22**, 4382 (1989).
74. P. Wisian-Neilson and M. A. Schaefer, *Macromolecules* **22**, 2003 (1989).
75. P. Wisian-Neilson and M. S. Islam, *Macromolecules* **22**, 2026 (1989).
76. R. H. Neilson and co-workers, *Macromolecules* **20**, 910 (1987).
77. H. R. Allcock and co-workers, *Chem. Mater.* **2**, 425 (1990).
78. F. J. Garcia-Alonso and P. Wisian-Neilson, *Polym. Preprints* **34**(1), 264 (1993).
79. H. R. Allcock, S. M. Coley, I. Manners, G. Renner, and O. Nuyken, *Macromolecules* **24**, 2024 (1991).
80. J. A. Dodge, I. Manners, H. R. Allcock, G. Renner, and O. Nuyken, *J. Am. Chem. Soc.* **112**, 1268 (1990).
81. Y. Z. Ni and co-workers, *Macromolecules* **25**, 7119 (1992).
82. H. W. Roesky and M. Lücke, *Angew Chem. Int. Ed. Eng.* **28**, 493 (1989).
83. A. K. Roy, *J. Am. Chem. Soc.* **115**, 2598 (1993).
84. A. K. Roy, G. T. Burns, G. C. Lie, and S. Grigoras, *J. Am. Chem. Soc.* **115**, 2604 (1993).

85. U.S. Pat. **5,068,379** (Nov. 26, 1991), A. K. Roy (to Dow Corning Corp.); U.S. Pat. **5,194,557** (Mar. 16, 1993), A. K. Roy and G. T. Burns (to Dow Corning Corp.).

86. S. Yajima, J. Hayashi, and M. Omori, *Chem. Lett.*, 931 (1975).

87. U.S. Pat. **4,159,259** (1979), J. Hayashi, M. Omori, and S. Yajima.

88. Y. Hasegawa, M. Imura, and S. Yajima, *J. Mater. Sci.* **15**, 720 (1980).

89. S. Yajima, *Ceram. Bull.* **62**, 993 (1983).

90. J. P. Wesson and T. C. Williams, *J. Polym. Sci. Polym. Chem. Ed.* **18**, 959 (1980).

91. R. West and co-workers, *J. Am. Chem. Soc.* **103**, 7352 (1981).

92. U.S. Pat. **4,260,780** (1981), R. West.

93. R. E. Trujillo, *J. Organomet. Chem.* **198**, C27 (1980).

94. R. D. Miller and J. Michl, *Chem. Rev.* **89**, 1359 (1989).

95. S. H. Yi, N. Maeda, T. Suzuki, and H. Sato, *Polymer J. (Tokyo)* **24**, 865 (1992).

96. M. Fujino, T. Hisaki, M. Fujiki, and N. Matsumoto, *Macromolecules* **25**, 1079 (1992).

97. H. K. Kim, C. R. Hove, and C. K. Ober, *J. Macromol. Sci. Pure Appl. Chem.* **A29**, 787 (1992).

98. R. West, *J. Organomet. Chem.* **300**, 327 (1986).

99. R. H. Cragg, R. G. Jones, A. C. Swain, and S. J. Webb, *J. Chem. Soc., Chem. Comm.* 1147 (1990).

100. S. Gauthier and D. J. Worsfold, *Macromolecules* **22**, 2213 (1989).

101. R. D. Miller, E. J. Ginsburg, P. Jenkner, and D. Thompson, *Polym. Preprints* **34**(1), 232 (1993).

102. G. J. Price, *J. Chem. Soc. Chem. Comm.* 1209 (1992).

103. R. G. Jones, R. E. Benfield, R. H. Cragg, and A. C. Swain, *J. Chem. Soc. Chem. Comm.*, 112 (1992).

104. J. M. Zeigler, L. I. McLaughlin, and R. J. Perry, *J. Inorg. Organomet. Polym.* **1**, 531 (1991).

105. J. F. Harrod, in Ref. 14, Chapt. 7, p. 89.

106. Y. Mu, C. Aitken, B. Cote, and J. F. Harrod, *Can. J. Chem.* **69**, 264 (1991).

107. H.-G. Woo and T. D. Tilley, in D. R. Uhlmann and D. R. Ulrich, eds., *Ultrastructure Processing of Advanced Materials*, John Wiley & Sons, Inc., New York, 1992, p. 651.

108. J. Y. Corey and X-H. Zhu, *J. Organomet. Chem.* **439**, 1 (1992).

109. M. Cypriak, Y. Gupta, and K. Matyjaszewski, *J. Am. Chem. Soc.* **113**, 1046 (1991).

110. J. Chrusciel, E. Fossum, and K. Matyjaszewski, *Polym. Preprints*, **34**(1), 221 (1993).

111. K. Sakamoto, K. Obata, H. Hirata, M. Nakajima, and H. Sakurai, *J. Am. Chem. Soc.* **111**, 7641 (1991).

112. H. Sakurai, K. Sakamoto, Y. Funada, and M. Yoshida, *Polym. Preprints* **34**(1), 218 (1993).

113. C. Biran, M. Bordeau, P. Pons, M. Leger, and J. Dunogues, *J. Organomet. Chem.* **382**, C17 (1990).

114. T. Shono, S. Kashimura, M. Ishifune, and R. Nishida, *J. Chem. Soc., Chem. Comm.* 1160 (1990).

115. K. Matyjaszewski and co-workers, *J. Macromol. Sci. Chem.* **A28**, 1151 (1991).

116. K. Matyjaszewski, *J. Inorg. Organomet. Polym.* **1**, 463 (1991).

117. J. P. Banovetz, Y.-L. Hsiao, and R. M. Waymouth, *Polym. Preprints* **34**(1), 228 (1993).

118. R. D. Miller and R. Sooriyakumaran, *J. Polym. Sci. Polym. Lett. Ed.* **25**, 321 (1987).

119. R. D. Miller and R. Sooriyakumaran, *Macromolecules* **21**, 3120 (1988).

120. P. Trefonas III and R. West, *J. Polym. Sci. Polym. Lett. Ed.* **23**, 469 (1985).

121. F. C. Schilling and co-workers, *Macromolecules* **22**, 3055 (1989).

122. F. C. Schilling and co-workers, *J. Inorg. Organomet. Polym.* **2**, 47 (1992).

123. H. S. Plitt and J. Michl, *Chem. Phys. Lett.* **198**, 400 (1992).

124. R. West and co-workers, *J. Am. Chem. Soc.* **103**, 7352 (1981).

125. R. G. Kepler, J. M. Zeigler, L. A. Harrah, and S. R. Kurtz, *Phys. Rev. B* **35**, 2818 (1987).

126. M. Fujino, *Chem. Phys. Lett.* **136**, 451 (1987).

127. M. Stolka, H.-J. Yuh, K. McGrane, D. M. Pai, *J. Polym. Sci. Polym. Chem. Ed.* **25**, 823 (1987).

128. C. L. Callender, *J. Opt. Soc. Am. B.* **9**, 518 (1992).

129. R. D. Miller and co-workers, in Ref. 14, p. 43.

130. R. Baney and G. Chandra, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, 2nd ed., Vol. 13, John Wiley & Sons, Inc., New York, 1988, p. 312.

131. G. M. Wallraff, R. D. Miller, M. Baier, E. J. Ginsburg, and R. R. Kunz, *J. Photopolym. Sci. Technol.* **5**, 111 (1992).

132. R. R. Kunz and M. W. Horn, *Polym. Eng. Sci.* **32**, 1595 (1992).

133. Y. Yamaguchi, Y. Hiroshige, M. Kakui, and M. Yokoyama, *J. Imaging Sci. Technol.* **36**, 256 (1992).

134. Eur. Pat. **0 475 349 A1** (Dec., 1992), M. Yokoyama and M. Kakui (to Mita Industrial Co., Ltd.).

135. R. West, A. R. Wolff, and D. J. Peterson, *J. Radiat. Curing* **13**, 35 (1986).

136. A. R. Wolff and R. West, *Appl. Organomet. Chem.* **1**, 7 (1987).

137. P. Trefonas and R. West, *J. Polym. Sci. Polym. Chem. Ed.* **23**, 2099 (1985).

138. R. D. Miller and R. Sooriyakumaran, *J. Polym. Sci. Polym. Chem. Ed.* **25**, 111 (1987).

139. H. Isaka, M. Fujiki, M. Fujino, and N. Matsumoto, *Macromolecules* **24**, 2647 (1991).

140. T. Shono, S. Kashimura, and H. Murase, *J. Chem. Soc. Chem. Comm.*, 896 (1992).

141. K. Takeda, K. Shiraishi, and N. Matsumoto, *J. Am. Chem. Soc.* **112**, 5043 (1990).

142. H. Tachibana and Y. Kawabata, *Phys. Rev. B.* **45**, 8752 (1992).

143. J.-C. Baumert and co-workers, *Appl. Phys. Lett.* **53**, 1147 (1988).

General References

J. M. Zeigler and F. W. G. Fearon, eds., *Silicon-Based Polymer Science*, Advances in Chemistry Series, Vol. 224, American Chemical Society, Washington, D.C., 1990; regarding poly(silanes).

M. Zeldin, K. J. Wynne, and H. R. Allcock, eds., *Inorganic and Organometallic Polymers*, ACS Symposium Series, Vol. 360, American Chemical Society, Washington, D.C., 1988; regarding poly(phosphazenes), poly(silanes), and other polymers.

Inorganic and Organometallic Polymers II: Advanced Materials and Intermediates, ACS Symposium Series, Vol. 572, American Chemical Society, Washington, D.C., 1994; regarding inorganic polymers.

Aroop K. Roy
Dow Corning Corporation

INORGANIC REFRACTORY FIBERS.

See FIBERS, ACRYLIC; REFRACTORY FIBERS.

INOSITOL.

See VITAMINS.

INSECT CONTROL TECHNOLOGY

Insects are the most numerous of living organisms, and the nearly one million described species constitute approximately 72% of all animal species (1). Of these about 1% are considered significant pests; they attack humans and or domestic animals, transmit human, animal, and plant diseases, destroy structures, and compete for available supplies of food and fiber. In the United States there are more than 10,000 species of insects, mites, and ticks that cause losses to agriculture, but only about 600 species require annual applied control measures. About 40% of these (235 species of 600) are exotic importations, illustrating the vulnerability of U.S. agriculture to infestations resulting from global travel (1).

Careful estimates suggest that the total losses to agricultural crops from insect attacks in the United States average about 10% of production and amount to more than \$14 billion annually (1). Worldwide agricultural losses from insect attacks have been estimated as about 14% of production (2).

Termites may cause more direct monetary damage than any other group of insect pests. It has been estimated that termites damage human made structures annually to the extent of 1% of their value in the United States and to 10% in the tropics.

Losses resulting from the depredations by insect vectors of human and animal diseases are almost beyond monetary estimation (1,3). Malaria transmitted by the bites of some 60 species of *Anopheles* mosquitoes, results from the infection of human erythrocytes by four species of protozoa, *Plasmodium falciparum*, *P. malariae*, *P. ovale*, and *P. vivax*. Despite a global program of eradication and control, begun in 1955, the World Health Organization (WHO) estimated in 1990 that there were annually about 270 million cases of malaria with about one million deaths. The number of persons suffering from lymphatic filariasis transmitted largely by the common house mosquito *Culex pipiens* and caused by the nematode parasites *Wuchereria bancrofti* and *Brugia malayi* is estimated at 100 to 250 million annually. The number of persons at risk from this predominantly urban disease has doubled over the past several decades. Trachoma, a viral disease causing blindness, is widespread in India and North Africa. It is transmitted by the housefly *Musca domestica*, and an estimated 80 million persons are infected. African trypanosomiasis (sleeping sickness) is caused by *Trypanosoma gambiense* and *T. rhodesiense* and is transmitted by the biting of several species of tsetse flies, *Glossina* spp. This disease and the related cattle disease (nagana), caused by *T. brucei*, have retarded the economic development of some 11 million km² of equatorial Africa. A number of species of biting black flies, *Simulium* spp., transmit the filarian *Onchocerca volvulus* to humans. This disease afflicts an estimated 20 million humans in the Volta River basin of Africa and in Central America. It is the cause of river blindness which has greatly retarded agricultural and economic development in the infested areas. American trypanosomiasis is caused by *Trypanosoma cruzi*, which is transmitted by a number of species of Triatomini "kissing bugs." Approximately 10 to 12 million humans in South and Central America are afflicted with this disease.

Other human diseases transmitted by insects include plague, *Yersina pestis*, caused by the bites of the oriental rat flea *Xenopsylla cheopis* and other fleas, and epidemic typhus, *Rickettsia prowazekii*, transmitted by the bite of the human body louse, *Pediculus humanus*. Insects are the vectors of more than 250 viruses (arboviruses) that are pathogens of humans and higher animals. These include yellow fever and dengue hemorrhagic fever for which the mosquitoes *Aedes aegypti* and *A. albopictus* are the important vectors. Human encephalitides such as the California, St. Louis, eastern, western, and Venezuelan types, are transmitted by the biting of a variety of *Aedes* and *Culex* mosquitoes (1).

Role of Chemicals in Insect Control. Plant-derived insecticides, eg, nicotine, rotenone, veratrine, and pyrethrum, have been used to kill insect pests since antiquity. A tea of nicotine from tobacco leaves was recommended to control aphids in 1793, but the principal development of insecticides for crop protection began about 1865 with the use of paris green, an arsenical stomach poison, for the control of the Colorado potato beetle, *Leptinotarsa decimlineata*. An improved compound, lead arsenate, was introduced in 1892 and calcium arsenate in 1907; their combined production in the United States, approximately 40 million kg annually, was applied predominately for the control of cotton insects. Cryolite was introduced as a stomach poison insecticide in 1928, to avoid objectionable arsenical residues on fruits and vegetables. These stomach poison insecticides were effective only against chewing insects and had little or no contact action. The arsenicals had the grave disadvantages of high toxicity to humans and domestic animals, considerable phytotoxicity, and extreme environmental persistence.

The first practical synthetic organic insecticide was the potassium salt of 4,6-dinitro-2-methylphenol [534-51-1] developed in Germany in 1892 as a dormant spray for orchard pests. However, it was the discovery of the insecticidal properties of DDT in 1939 that began an era of chemical pest control resulting in the synthesis and evaluation of hundreds of thousands of synthetic organic chemicals as insecticides (4). Dichlorodiphenyltrichloroethane (DDT), with its efficient contact insecticidal action together with long residual persistence and relative safety to humans and domestic animals, largely replaced the arsenicals. Hundreds of new uses were developed and DDT production in the United States attained a maximum of 77,800 t in 1961. Massive use of other organochlorines, such as benzene hexachloride, chlorinated camphenes, chlordane, heptachlor, aldrin, dieldrin, and endrin followed, and by 1964 these chemicals represented 70% of the total (53,000 t) use of insecticides in agriculture together with organophosphates at 20% (10,600 t). However, as insecticide resistance supervened and concern over environmental pollution increased, the use of the organophosphates (introduced in 1948), and that of the carbamates (introduced in 1957) increased rapidly. By 1976, more than 200 chemical compounds were marketed as insecticides and the total application to primary crops was 58,000 t. The organochlorines represented 29%, the organophosphates 49%, and the carbamates 19%. Cotton was the most heavily treated crop (49% of the total) followed by corn, 25%; and soybean, 6%. Increasing use of integrated pest management (IPM) practices and the introduction of the pyrethroids, which are effective at about one-tenth the application rate of the older insecticides, resulted in decreased insecticide use and by 1982 the farm use on primary crops was estimated at 32,000 t, comprised of organochlorines, 6%; organophosphates, 67%; carbamates, 18%; and pyrethroids, 4%. Corn became the most heavily treated crop with 42% of the total, followed by cotton, 24%, and soybean, 16%.

Since the early 1940s, insecticides have been of immeasurable value in curbing the ravages of insect pests. In the words of the National Academy of Sciences "...when their use is approached from sound ecological principles, chemical pesticides provide dependable and valuable tools for the biologist. Their use is indispensable in modern society. There are many problems of insect pest control for which the use of chemicals provides the only acceptable solution. Chemical pesticides will continue to be one of the most dependable weapons for the entomologist for the foreseeable future" (6).

In agriculture, the average benefit cost ratio from insecticide use ranges from \$3 to \$5 return for every \$1 invested by the farmer(s). There are many examples where the return is much greater. In California, treatment of sugarbeets with granular phorate systemic insecticide to control the aphid and

leafhopper vectors of virus yellows increased sugar yields up to 1685 kg ha (1500 lb acre), for a ratio of \$18 to \$1 invested. The use of DDT in Wisconsin to control the Colorado potato beetle and the potato leafhopper, *Empoasca fabae*, increased yields by as much as 5.7 m³ ha (65 bushel acre) for a ratio of \$29 to \$1. It is scarcely possible to produce apples, sweet corn, lettuce, or broccoli of modern marketable quality without the use of insecticides.

The value of insecticides in controlling human and animal diseases spread by insects has been dramatic. It has been shown that between 1942 and 1952, the use of DDT in public health measures to control the mosquito vectors of malaria and the human body louse vector of typhus saved five million lives and prevented 100 million illnesses (4). Insecticides have provided the means to control such important human diseases as filariasis transmitted by *Culex* mosquitoes and onchocerciasis transmitted by *Simulium* blackflies.

Integrated Pest Management. Although employment of chemicals for insect pest control is essential to modern society, the extensive and injudicious use of chemical insecticides since 1946 has resulted in many problems including (1) widespread insect resistance, (2) emergence of resurgent and secondary pests whose regulating natural enemies are adversely affected, (3) hazards to human health, (4) ubiquitous environmental pollution by persistent lipophilic organochlorines, and (5) exponentially increasing costs of new insecticides (6–10). Most of these unintended consequences of chemical pest control relate to a pervasive eradication philosophy resulting from the euphoria about the effectiveness of successive generations of organochlorine, organophosphorus, carbamate, and pyrethroid insecticides. WHO initiated a global eradication program for malaria based on annual residual house spraying directed at the *Anopheles mosquito* vectors. In the United States, eradication programs by insecticide applications were undertaken against such pests as the gypsy moth *Lymantria dispar*, the Japanese beetle *Popillia japonica*, the red imported fire ant *Solenopsis invicta*, and the yellow fever mosquito *Aedes aegypti*. Initial reaction was to exploit yet another group of insecticides with different chemistry as each program faltered because of insecticide resistance, destruction of wildlife, and environmental pollution. Thus humans are confronted by the anomaly that since the 1940s, despite a 10-fold increase in the use of insecticides and a 20-fold proliferation of the chemicals available for insect control, there has been little change in the benefit cost ratio to the farmer. Repeated estimates by the U.S. Department of Agriculture have not indicated any appreciable decrease in the extent of insect damage to the crops most heavily treated with insecticides, ie, alfalfa, apple, corn, cabbage, cotton, and potato. Average losses due to insect and mite pests in these crops were estimated as follows: 1900–1904, 11.3° o; 1910–1935, 14.0° o; 1942–1951, 11.0° o; and 1951–1960, 13.3° o (7).

It is apparent that insect pest control must be directed away from exclusive reliance on insecticides and toward the optimization of pest control tactics in an ecologically and economically sound way. Integrated pest management (IPM) has been variously defined as (1) "a system in which all available techniques are evaluated and consolidated into a unified program to regulate pest populations so that economic damage is avoided and environmental disturbances are minimized," (2) as "the intelligent selection of and use of pest control actions that will ensure favorable economic, ecological, and sociological consequences," or (3) as "the selection, integration, and implementation of pest control based on predicted economic, ecological, and sociological consequences" (6,10).

The primary goals of IPM are (1) to determine how the life system of the pest needs to be modified to reduce the numbers to tolerable levels, ie, below the economic threshold; (2) to apply biological knowledge and current technology to achieve the desired modification, ie, applied ecology; and (3) to devise procedures for pest control compatible with economic and environmental control aspects, ie, economic and social acceptance (9).

IPM practices rely heavily on protection and conservation of natural enemies, parasites, predators, and diseases that regulate or balance populations of insect pests, thus IPM rejects the regular or preventive use of broad-spectrum insecticides and the general philosophy of pest species eradication, which has proved unworkable. IPM programs are based on two important parameters: the economic injury level defined as that population density of a pest that causes enough injury to justify the cost of remedial treatment, and the economic threshold, defined as that pest density at which control measures should be applied to prevent an increasing insect population from reaching the economic injury level. Thus IPM is a dynamic concept akin to game, fisheries, and forest management.

Whenever applied, IPM practices have consistently resulted in decreases in insecticide applications of 50 to 90° o over conventional spray programs. By encouraging natural enemies, IPM practices markedly decrease the rigor of natural selection by pesticides that is responsible for resistance. Natural enemy preservation also prevents the great fluctuations and surges in insect pest populations observed after the injudicious use of broad-spectrum insecticides. Under the IPM concept, insecticides are generally used when other practices are inadequate and the pest population reaches the economic threshold. In order to make the IPM concept effective, insecticides must be used as selectively as possible, with minimal disturbance to all other elements of the ecosystem. Thus IPM practices are essentially blueprints for the proper use of insecticides in insect pest control.

Insecticide management is concerned with the safe, efficient, and economical handling of insecticides during manufacture, utilization, and disposal. The essential components are selection of the proper insecticide for the IPM program, selection of the mode, timing, and dosage of application, consideration of the problems of resistance and resurgence, the possible effects of insecticide residues on food crops, and in the environment, and the impact of these on humans, domestic animals, and wildlife (7,9).

This article provides general information on many aspects of the chemical control of insect pests as developed during the twentieth century. It summarizes the chemistry, properties, uses, and advantages and disadvantages of most of the chemicals used for insect control, including many products of largely historical interest as well as some not registered for use in the United States but used elsewhere in the world (11–18).

Insecticides

Inorganic Stomach Poisons.

Arsenicals. Various arsenical salts have been widely used as stomach poisons (19) for chewing insects. Arsenious oxide, [1327-53-3] As₂O₃, secured from flue dust after the roasting of various metallic ores, is the principal source of these arsenical insecticides, and forms arsenous acid, H₃AsO₃, a weak monobasic acid, *K*₁ = 6.0 × 10⁻¹⁰. Copper acetate arsenite or paris green, Cu(CH₃COO)₂ · 3Cu(AsO₂)₂, was the first synthetic insecticide, discovered in about 1865 for the control of the Colorado potato beetle. The rat oral LD₅₀ is 22 mg/kg. Sodium arsenite, NaAsO₂, rat oral LD₅₀ 10–50 mg/kg, has been used extensively in poison baits for grasshopper control and as a dip to control livestock ectoparasites. Arsenic pentoxide, As₂O₅, forms tribasic orthoarsenic acid, *K*₁ = 2.5 × 10⁻⁴, *K*₂ = 5.6 × 10⁻⁸, and *K*₃ = 3 × 10⁻¹³. Acid lead arsenate, PbHAsO₄, rat oral LD₅₀ 800 mg/kg, water-soluble to 2.5 g/L, was developed in 1892 for the control of the gypsy moth. It has become the principal insecticide for the control of the codling moth which attacks deciduous fruits. Basic lead arsenate, Pb₄(PbOH)(AsO₄)₃, is a safer form for use on plant foliage as it does not readily form arsenic acid. Calcium arsenate, Ca₃(AsO₄)₂, rat oral LD₅₀ 40–100 mg/kg, was found to be effective as undiluted dust for the control of the boll weevil in 1920. It became the most widely used insecticide during the first half of the twentieth century. U.S. production in 1942 was 38 million kg.

Mode of Action. The fundamental biochemical lesion produced by arsenicals is the result of reaction between As³⁺ and the sulfhydryl groups of key respiratory enzymes such as pyruvate and α-ketoglutarate dehydrogenases.

Fluorides. Problems with persistent residues of lead arsenate on deciduous fruits encouraged a search for safer stomach poison insecticides and led to the development in 1928 of cryolite or sodium fluoroaluminate, NaAlF₆. Cryolite occurs naturally as a mineral and is produced synthetically as a lighter, fluffier product. The rat oral LD₅₀ is about 13,500 mg/kg and it is soluble in water to 0.3 g/L. Cryolite was used extensively as a safe insecticide for chewing insects such as the codling moth, tomato and cabbage worms, flea beetles, and the Mexican bean beetle. Seven thousand metric tons were applied in the United States in 1944. It is still useful to control chewing insects without damaging important parasites and predators. Sodium fluoride, NaF, has been used as an insecticide since 1896 for the control of cockroaches and chewing lice (Mallophaga). The rat oral LD₅₀ is 200 mg/kg. It is water soluble to 43 g/L. Sodium fluorosilicate, Na₂SiF₆, is used for mothproofing (see REPELLENTS). The rat oral LD₅₀ is 125 mg/kg. It is water soluble to 6.5 g/L.

Mode of Action. The fluoride ion inhibits enzymes, such as enolase, which require Mg as a prosthetic group, by precipitating a complex magnesium fluorophosphate; thus it prevents phosphate transfer in oxidative metabolism.

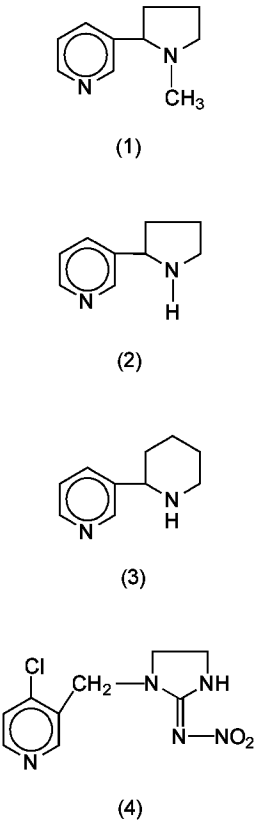
Miscellaneous Inorganic Insecticides. Many inorganic substances have had limited usefulness in insect control but have been superseded by much safer organic insecticides. Borax, Na₂B₄O₇ · 10H₂O, and sodium tetraborate, Na₂B₄O₇, have been used to kill housefly maggots in manure or refuse, to prevent the breeding of mosquito larvae in water to be used only for laundering, or as glyceroboric acid to treat maggot-infested wounds in animals. Boric acid, H₃BO₃, has been used as a stomach poison for cockroaches (see also BORON COMPOUNDS, BORIC ACID).

White phosphorus, incorporated in sweet syrup, forms a useful bait for cockroaches. Silicic acid, SiO₂ or H₂SiO₃, very finely divided, is a rapidly acting desiccant that kills cockroaches, fleas, termites, and stored-grain pests by dehydration.

Sulfur and its compounds are among the oldest and most widely used pesticides. Elemental sulfur is especially effective as a dust for the control of mites attacking citrus, cotton, and field crops and as a protectant against chiggers, *Trambicula* spp., attacking humans. Sulfur also is a valuable fungicidal diluent for other dust insecticides and is used in wettable form as a spray mixture. Lime sulfur has been a standard dormant spray for the control of the San Jose scale, *Quadraspidiotus perniciosus*, and for other scales and various plant diseases. Lime sulfur is a water-soluble mixture of calcium pentasulfide, CaS₅, calcium tetrasulfide, CaS₄, and calcium thiosulfate, CaS₂O₃, that form with calcium sulfite, CaSO₃, which is relatively water insoluble, when lump or stone lime (1 part by wt), ground sulfur (2 parts), and water (ca 8 parts) are boiled together. The filtered stock material, which is a deep orange, malodorous liquid (sp gr 1.28) is diluted by ca 1 part to 7 or 8 parts of water (to sp gr 1.035) for dormant winter spraying and by ca 1 part to 49 parts of water (to sp gr 1.005) for summer spraying. Dehydrated dry lime sulfur is available commercially and is used after appropriate dilution with water. A typical preparation contains 65–70% CaS₄ and CaS₅, ca 5% CaS₂O₃, 5–10% free sulfur, and ca 20% inert ingredients.

Contact Poisons of Plant Origin.

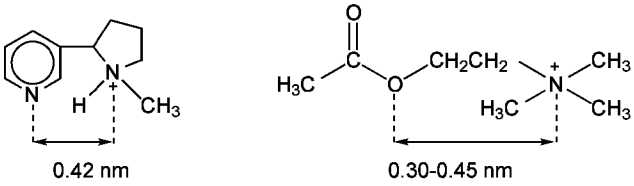
Nicotinoids. Nicotine from tobacco was one of the earliest insecticides and was recommended for use in 1763 as a tea for the destruction of aphids (1,20). Nicotine [54-11-3], L-1-methyl-2-(3'-pyridyl)pyrrolidine (1) (bp 247°C, d 1.009), is found in the leaves of *Nicotiana tobacum* and *N. rustica* (Solanaceae) in amounts ranging from 2 to 14%, and also is found in *Duboisia hopwoodii* and in *Aesclepias syriaca*. It occurs as the principal alkaloid along with small amounts of 12 other alkaloids of which normicotine [494-97-3], 2-(3'-pyridyl)pyrrolidine (2) (bp 270°C, d 1.07 g mL), and anabasine [494-52-0], L-2-(3'-pyridyl)piperidine (3) (bp 281°C, d 1.048), are of insecticidal importance (see ALKALOIDS). Normicotine occurs as both the D and L forms, the former in *D. hopwoodii* and the latter commonly predominating in *Nicotiana*. Anabasine is the chief alkaloid of *Anabasis aphylla*, where it occurs from 1–2% in the shoots and is found to ca 1% in *Nicotiana glauca*.



These nicotinoids are appreciably volatile (nicotine vapor pressure, 5.7 Pa at 25°C) and, although colorless liquids when pure, rapidly darken upon exposure to air. They are highly basic ($K_{b1} = 1 \times 10^{-6}$, $K_{b2} = 1 \times 10^{-11}$) and readily form salts with acids and many metals. Nicotine sulfate [65-30-5], (C₁₀H₁₄N₂)₂·H₂SO₄, is preferred as an insecticide because it is more stable and less volatile. Nicotine has an oral LD₅₀ to the rat of 30 mg/kg and a dermal LD₅₀ to the rabbit of 50 mg/kg.

Imidocloprid [105827-78-9] 1-[(6-chloro-3-pyridinyl)methyl]-N-nitro-2-imidazolidinimine (4) (bp 137–144°C, vp 0.2 μPa at 20°C) is soluble to 0.51 g/L. It is a synthetic nicotinoid with both contact and systematic activity against aphids, leafhoppers, whiteflies, and other sucking insects. The rat LD₅₀s are 424 male, 475 female (oral), and 5000 (dermal) mg/kg.

Mode of Action. Nicotine, anabasine, and imidocloprid affect the ganglia of the insect central nervous system, facilitating transsynaptic conduction at low concentrations and blocking conduction at higher levels. The extent of ionization of the nicotinoids plays an important role in both their penetration through the ionic barrier of the nerve sheath to the site of action and in their interaction with the site of action, which is believed to be the acetylcholine receptor protein. There is a marked similarity in dimensions between acetylcholine and the nicotinium ion.

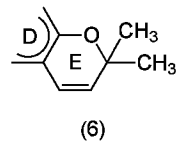
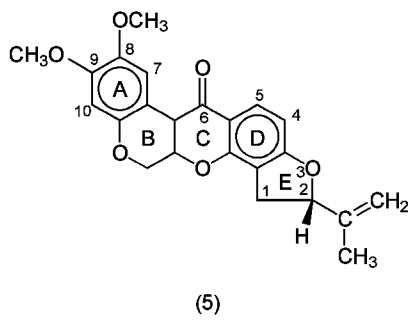


Thus nicotinoids that have the highest insecticidal action have the highest pK_a and, consequently, exist largely in the ionized form at physiological pH. This produces the anomaly that the compounds that are most highly ionized react most rapidly with the receptor protein, yet they are less able to penetrate through the ionic barrier surrounding the insect nerve synapse.

Nicotine is used as a contact insecticide for aphids attacking fruits, vegetables, and ornamentals, and as a fumigant for greenhouse plants and poultry mites. Nicotine sulfate is safer and more convenient to handle and the free alkaloid is rapidly liberated by the addition of soap, hydrated lime, or ammonium hydroxide to the spray solution. Nicotine sprays commonly contain 0.05–0.06% nicotine, and nicotine dusts, 1–2% nicotine.

Rotenoids. The use of rotenone-bearing roots as insecticides in the United States was developed as a result of federal laws against residues of lead, arsenic, and fluorine upon edible produce. Rotenone [83-79-4] (5) is harmless to plants, highly toxic to many insects, and relatively innocuous to

mammals. Out of several species of plants, including 21 species of *Tephrosia*, 12 *Derris*, 12 *Lonchocarpus*, 10 of *Milletia*, and several of *Mundulea*, 68 have been reported to contain rotenone or rotenoids. The principal economic species are the various kinds of derris, *Derris elliptica* and *D. malaccensis*, from Malaya and the Indonesian islands, and cubü, or timbo, *Lonchocarpus utilis* and *L. urucu*, from South America. *Tephrosia (Cracca) virginiana* has been studied as a possible native source of rotenone. A number of toxic constituents have been isolated from the roots and seeds of these plants, the most important of which is rotenone (mp 163°C (a dimorphic form exists, mp 181°C)).



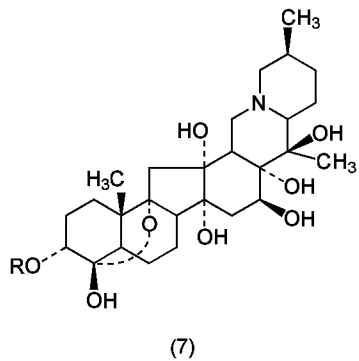
Other naturally occurring rotenoids are elliptone [478-10-4] (mp 159°C), which has a furan ring in place of ring E (double bond between carbons 1 and 2); sumatrol [82-10-0] (mp 188°C), which is 5-hydroxyrotenone; malaccol [478-07-9] (mp 244°C), which is 5-hydroxyelliptone; 1- α -toxicarol [82-09-7] (mp 101°C), which has a hydroxy group at carbon 5 and the ring (6) in place of ring E of rotenone; and deguelin [522-17-8] (mp 165–171°C), which has a hydrogen atom on carbon 5 in place of the hydroxyl group of toxicarol.

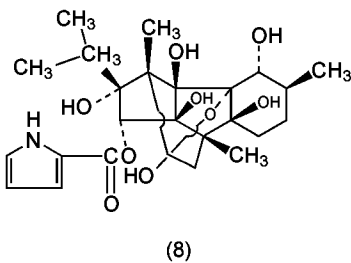
A related material, tephrosin [76-80-2] (mp 197–198°C), which has a hydroxyl group on one of the carbon atoms between rings B and C, does not appear to occur naturally in derris resins but is an oxidation product of deguelin. All the naturally occurring rotenoids exist as levo forms. Rotenone is from five to ten times as effective insecticidally as the other rotenoids. The content of these materials in commercial plant species is variable: roots of *D. elliptica* averaging from 5–9% rotenone (max 13%) with up to 31% ether extractives; *D. malaccensis* contains up to 4% rotenone and up to 27% extractives. *L. utilis* averages 8–11% rotenone as a max and 25% total extractives. Toxicarol and deguelin make up the bulk of the total extractives. When exposed to light and air, rotenone decomposes, changing from colorless through yellow to deep red with the formation of a hydroxyl derivative at C-7 and then dehydration to form dehydrorotenone with a C=C bond between rings B and C. These reactions render rotenone residues inactive after 5–10 d exposure to normal sunlight. Decomposition to dehydrorotenone is accelerated by the presence of alkali, and rotenone should be considered incompatible with alkaline dusts, eg, lime, or with soaps and other alkaline wetting and spreading agents. Pure crystalline rotenone is prepared by extracting powdered rotenone-containing roots with a solvent, eg, ether or carbon tetrachloride, and concentrating the solution to produce crystallization. The pure rotenone, rat oral LD₅₀ 130 mg kg, is soluble in water to 20 μ g L.

Insects poisoned with rotenone exhibit a steady decline in oxygen consumption and the insecticide has been shown to have a specific action in interfering with the electron transport involved in the oxidation of reduced nicotinamide adenine dinucleotide (NADH) to nicotinamide adenine dinucleotide (NAD) by cytochrome *b*. Poisoning, therefore, inhibits the mitochondrial oxidation of Krebs-cycle intermediates which is catalyzed by NAD.

Rotenone-containing insecticides have been used as dusts of ground roots, dispersible powders, and emulsive extracts. Their principal uses have been for application to edible produce just prior to harvest and for the control of animal ectoparasites and cattle grubs.

Sabadilla. The seeds of sabadilla, or *Schoenocaulon officinale*, family Liliaceae, of South and Central America have been used in native louse powders for centuries and have attained some commercial importance as an insecticide. The related species, *S. drummondii* and *S. texanum*, which are indigenous to the United States, also contain the toxic principles which are a complex group of alkaloids called veratrine. These are present in sabadilla seed to 2–4%, and include about 13% cevadine [62-59-9] (mp 205°C), 10% veratridine [71-62-5] (mp 160–180°C), and lesser amounts of cevadilline [1415-76-5], and sabadine [28903-30-2]. The crude alkaloids are soluble in water to 55 mg L at 4°C. Sabadilla preparations are markedly activated by heating to 75–80°C for 4 h, moistening with 10% sodium carbonate solution, or by milling with hydrated lime, which is most effective. The toxic alkaloids are rapidly destroyed by the action of light, thus sabadilla, which is both a contact and a stomach poison, is safe to use on food crops within a few days of harvest. Cevadine, C₃₂H₄₉O₉N (7), R=CH₃CH=CCH₃CO—, and veratridine, C₃₃H₅₁O₁₁N (7), R=3,4 dimethoxybenzoyl, both esters of cevine [124-98-1] (7), R=H, are the alkaloids responsible for the insecticidal action. These alkaloids are highly poisonous to mammals and may cause irritation of the eyes and respiratory tract and violent sneezing. The ground seeds have an oral LD₅₀ to the rat of 5000 mg kg. Sabadilla is used as a dust or wettable powder of the ground seeds for the control of plant-feeding Hemiptera, and with sugar as a toxic bait for thrips.





Ryania. The root and stem of the plant *Ryania speciosa*, family Flacourtiaceae, native to South America, contain from 0.16–0.2% of insecticidal components, the most important of which is the alkaloid ryanodine [15662-33-9], C₂₅H₃₅O₉N (8) (mp 219–220°C). This compound is effective as both a contact and a stomach poison. Ryanodine is soluble in water, methyl alcohol, and most organic solvents but not in petroleum oils. It is more stable to the action of air and light than pyrethrum or rotenone and has considerable residual action. Ryania has an oral LD₅₀ to the rat of 750 mg/kg. The material has shown considerable promise in the control of the European corn borer and codling moth and is used as a wettable powder of ground stems or as a methanolic extract. Ryanodine uncouples the ATP–ADP actomyosin cycle of striated muscle.

Pyrethroids. The flowers of *Chrysanthemum cinerariaefolium* and *C. coccineum* (Compositae) have been recognized as insecticidal since about 1800, and the manufacture of flea and louse powders began in the Transcaucasis region of Asia about 1828 (17,20,21). The nature and source of the product pyrethrum were published in 1851 and its use became worldwide. Pyrethrum was a principal product of Dalmatia until World War I, when Japan became the leading producer. Production began in Kenya about 1932 and that country became the principal supplier by 1940, with about 21,000 t produced in 1987. Kenya flowers average about 1.3% of toxic ingredients, with 90% of these found in the mature fully opened flower heads. Pyrethrum concentrates are made by extracting the flowers with petroleum ether, acetone, or acetic acid, and these extracts are purified from waxes by reextraction with nitromethane or methanol and by adsorption on activated charcoal to produce a 25% pyrethrum concentrate containing about 10% pyrethrins I, 9% pyrethrins II, 2% cinerin I, 3% cinerin II, 1% jasmolin I, and 1.1% jasmolin II. These active components consist of two cyclopropanecarboxylic acids, chrysanthemic acid [10453-89-1] (9), R = CH₃, and pyrethoic acid [26767-71-5], R = COOCH₃; esterified with three cyclopentenylketo alcohols (10), pyrethrolone [487-67-2], cinerolone [17190-74-8], and jasmololone [22054-39-3] (Table 1). The esters are asymmetric with absolute configurations of 1R,3R, and 4S, and the C=C bond in the acid side chain is trans and in the alcohol cis.

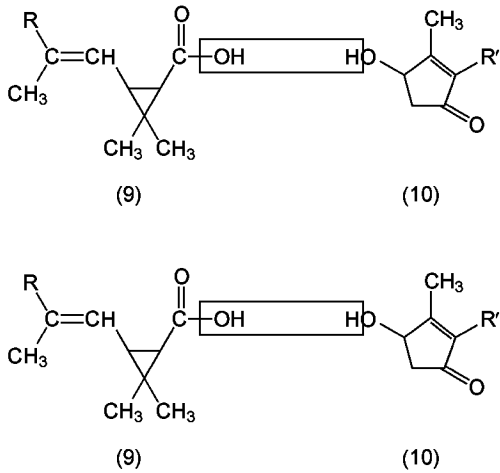


Table 1. Esters of Acid^a and Alcohol^b

Ester	CAS Registry Number	R in (9)	R' in (10)
pyrethrin I	[121-21-1]	CH ₃	CH ₂ CH=CHCH=CH ₂
pyrethrin II	[121-29-9]	COOCH ₃	CH ₂ CH=CHCH=CH ₂
cinerin I	[25402-06-6]	CH ₃	—CH ₂ CH=CHCH ₃
cinerin II	[121-20-0]	COOCH ₃	CH ₂ CH=CHCH ₃
jasmolin I	[4466-14-2]	CH ₃	CH ₂ CH=CHCH ₂ CH ₃
jasmolin II	[1172-63-0]	COOCH ₃	CH ₂ CH=CHCH ₂ CH ₃

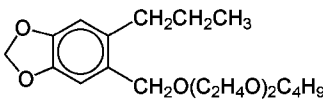
^a Structure (9).

^b Structure (10).

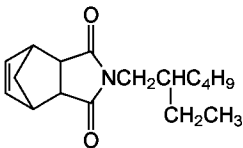
Pyrethrum is an especially valuable insecticide because it produces a rapid paralysis or knockdown of flying insects, and because its toxic components are rapidly inactivated upon exposure to light. The rat oral LD₅₀ values for pyrethrin I and cinerin I are approximately 1200 mg/kg. Pyrethrum products are used as household insecticides, livestock insecticides, grain protectants, and to control insect pests on edible produce just prior to harvest. The principal use is in aerosol sprays at 0.04 to 0.25% active ingredient (AI) together with 5–10 times this amount of a synergist such as piperonyl butoxide. **Synergists.** The pyrethrins and cinerins are rapidly detoxified in insects by microsomal oxidases incorporating cytochrome P450, that form epoxides with the C=C bonds of the side chains and carboxylic acids from terminal methyl groups; thus paralyzed insects often recover. Synergists containing the methylenedioxyphenyl moiety, although essentially nontoxic themselves, can activate the pyrethrins up to 30-fold by reacting with the Fe atom of the cytochrome to block detoxication. Such synergists are also highly effective when used with rotenone, ryania, and carbamates.

Piperonyl butoxide [51-03-6] is 5-[2-(2-butoxyethoxy)ethoxy]methyl-6-propyl-1,3-benzodioxole (11) (d 1.04–1.07, vp 0.13 kPa at 25°C). The rat oral LD₅₀s are 7500, 6150 mg/kg. Piperonyl butoxide is the synergist commonly used with natural pyrethrins in aerosol sprays.

MGK-264 [113-48-4], N-(2-ethylhexyl)-[2.2.1]-5-heptene-2,3-dicarboximide (12) (d 1.05) is an effective synergist of entirely different chemical structure (bp 158°C at 0.26 kPa). The rat oral LD₅₀ is 2800 mg/kg.



(11)



(12)

Synthetic Pyrethroid Insecticides. Elucidation of the chemical structures of the naturally occurring pyrethrum esters, their rapid and selective insecticidal action, and their high cost stimulated the search for effective synthetic derivatives (13,17,21). Since the 1940s, structural optimization has produced an array of broad-spectrum insecticides with activity 10- to 20-fold greater than other types of insecticides, and with extended residual action. These synthetic pyrethroids have become one of the most important classes of insecticides with world annual production estimated at 6000 t (21).

Pyrethroids from Chrysanthemic Acid. The unsaturated side chains of the allethrolone alcohol moieties of the natural pyrethrins are readily epoxidized by microsomal oxidases and converted to diols, thus detoxifying the insecticides. Esterification of chrysanthemic acid (9), R = CH₃, with substituted benzyl alcohols produces useful insecticides; barthrin [70-43-9], 2-chloro-3,4-methylenedioxybenzyl (±)-*cis,trans*-chrysanthemate, and dimethrin [70-38-2], 2,4-dimethylbenzyl (±)-*cis,trans*-chrysanthemate. These have a limited spectrum of insecticidal activity but are of very low mammalian toxicity, ie, rat oral LD₅₀s >20,000 mg/kg.

Allethrin [584-79-2] (*d* 1.005–1.015, vp 16 mPa at 30°C) is the allyl homologue of pyrethrin I, ie, R' = allyl, —CH₂CH=CH₂. The synthetic product contains 75–95% of eight enantiomers, 70% (±)-*trans* and 30% (±)-*cis* acids esterified with (±)-cyclopentenolone alcohol. The relative insecticidal activities of the enantiomers are shown in Table 2. The rat LD₅₀s of allethrin are 920 (oral) and 11,000 (dermal) mg/kg. Bioallethrin [584-79-2] is the (±)-*trans*-chrysanthemate ester of the (±)-alcohol (RS) (R) which has enhanced knockdown activity (vp 2 mPa at 25°C, water solubility 4.6 mg/L). Allethrin is as effective as the natural pyrethrins against flies and mosquitoes but has a narrow spectrum of activity against other insect pests.

Table 2. Allethrin Isomers with Their Abundance and Relative Toxicities to *Musca domestica*

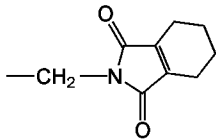
Allylrethronyl portion	Chrysanthemate portion	Percent	Relative toxicity ^a
L-	D- <i>trans</i> -	12.4	0.58
D-	L- <i>trans</i> -	12.4	0.14
D-	D- <i>trans</i> -	22.8	3.37
L-	L- <i>trans</i> -	22.8	0.02
L-	D- <i>cis</i> -	8.0	0.33
D-	L- <i>cis</i> -	8.0	0.14
D-	D- <i>cis</i> -	6.8	1.77
L-	L- <i>cis</i> -	6.8	0.06

^a The toxicity of DL-allylrethronyl DL-*cis,trans*-chrysanthemate is assumed to be equal to 1.00.

Resmethrin, (5-benzyl-3-furanyl)methyl (±)-*cis,trans*-chrysanthemate (13) (mp 43°C) is soluble in water to 0.3 mg/L. The product of four isomers contains 20–30% *cis* and 70–80% *trans* isomers (Table 3). The rat oral LD₅₀ is 2000 mg/kg. Bioresmethrin [28434-01-7], which is about twice as effective as resmethrin, is the corresponding (±)-*trans*-(1R,3R) chrysanthemate ester (mp 32°C). The rat LD₅₀s are >8000 (oral) and >1000 (dermal) mg/kg. Resmethrin is among the safest of all insecticides and is rapidly detoxified in light and air. It is used as a household and greenhouse insecticide.

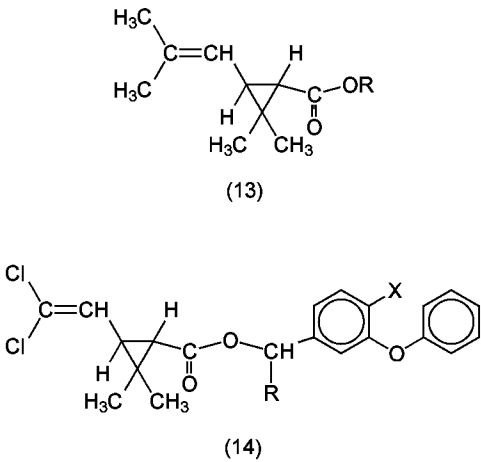
Table 3. Properties of *cis,trans*-Chrysanthemates^a

Name	CAS Registry Number	R	Density, ^b g/mL	Vapor pressure, ^b mPa ^c
resmethrin	[10453-86-8]			0.19
phenothrin	[26002-80-2]		1.058	0.16
cyphenothrin	[39515-40-7]		1.06	0.12
empenthrin	[54406-48-3]		0.934	8.7
prallethrin	[23031-36-9]		1.03	4.7

tetramethrin	[7696-12-0]		1.108	0.062
--------------	-------------	---	-------	-------

^a Structure (13).
^b At 20°C.
^c To convert mPa to mm Hg, multiply by 7.5 × 10⁻⁶

Phenothrin, (3-phenoxyphenyl)methyl (±)-*cis,trans*-chrysanthemate, a mixture of four isomers, has a rat oral LD₅₀ of >5000 mg/kg. Cyphenothrin is the α-CN-analogue of phenothrin (Table 3). It is a mixture of four isomers, and is soluble in water to 10 μg/L. The rat LD₅₀s are 318 male, 419 female (oral), and >5000 (dermal) mg/kg.



Empenthrin, (*E*)-(*R,S*)-1-ethynyl-2-methylpent-2-enyl (1*R,S*)-*cis,trans*-chrysanthemate, is soluble in water to 2.4 mg/L. The rat LD₅₀s are 2280 male, 1680 female (oral), and >5000 (dermal). Empenthrin is used for mothproofing (see also REPELLENTS).
Prallethrin, (*R,S*)-2-methyl-4-oxo-3-prop-2-ynylcyclopent-2-enyl(1*R,S*)-*cis,trans*-chrysanthemate, is soluble in water to 8.5 mg/L. The rat LD₅₀s are 640 male, 460 female (oral), and >5000 (dermal) mg/kg. Prallethrin is a household insecticide.

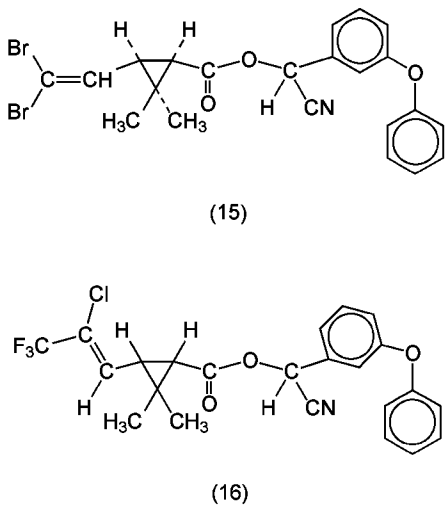
Tetramethrin, (1,3,4,5,6,7-hexahydro-1,3-dioxo-2*H*-isindole-2-yl)-methyl (±)-*cis,trans*-chrysanthemate (mp 65–80°C), a mixture of four isomers, is soluble in water to 20 μg/L. The rat oral LD₅₀ is >5000 mg/kg. Tetramethrin produces rapid knockdown and is used in veterinary hygiene.

Pyrethroids with Modified Chrysanthemate Esters. Newer pyrethroids incorporate optimized chrysanthemic acid components to retard detoxication by microsomal oxidases and these are esterified with a variety of optimized alcohol moieties therefore increasing persistence.

Permethrin [52645-53-1], (3-phenoxyphenyl)methyl-3-(2,2-dichloroethenyl)(±)-*cis,trans*-2,2-dimethylcyclopropanecarboxylate (14), R = H (mp 35°C, vp 45 μg/Pa at 25°C) is a mixture of four isomers of 70% (±)-*trans* and 30% (±)-*cis* esters. It is soluble in water to 0.2 mg/L. The rat oral LD₅₀s are 1500 and 2000 mg/kg. The incorporation of Cl atoms for the dimethylvinyl group substantially increases persistence. Thus permethrin is a broad-spectrum, persistent insecticide widely used for pests of cotton, corn, and other important crops. Cypermethrin [52315-07-8], (*R,S*)-α-cyano-(3-phenoxyphenyl)methyl (1*R,S*)-*cis,trans*-(2,2-dichloroethenyl)-2,2-dimethylcyclopropanecarboxylate (14), R = CN; X = H (mp 60–80°C, *d* 1.125, vp 0.51 mPa at 70°C) is a mixture of eight isomers of 70% (±)-*trans* and 30% (±)-*cis*. It is soluble in water to 0.01 mg/L. The rat LD₅₀s are 250, 415 (oral) and >1600 (dermal). Cypermethrin is more effective than permethrin and is widely used on agricultural crops and for cockroach control. α-Cypermethrin [67375-30-8] is the racemate of the (*S*), (1*R*) and (*R*), (1*S*) isomers.

Cyfluthrin [68359-37-5], (*R,S*)-α-cyano-(4-fluoro-3-phenoxyphenyl)methyl (1*R,S*)-*cis,trans*-3-(2,2-dichloroethenyl)-2,2-dimethylcyclopropanecarboxylate (14), R = CN; X = F (mp 60°C, vp <1 mPa at 20°C), a mixture of eight isomers, is soluble in water to 2 μg/L. The rat LD₅₀s are 500, 800 (oral) and 600 (dermal) mg/kg. Cyfluthrin is a broad-spectrum pyrethroid especially effective against greenhouse pests.

Deltamethrin [52918-63-5], (*S*)-α-cyano-3-phenoxyphenyl)methyl (+)-*cis*-(1*R*,3*R*) 3-(2,2-dibromoethenyl)-2,2-dimethylcyclopropanecarboxylate (15) (mp 98–101°C, *d* 1.108, vp 2 μPa at 25°C) is water soluble to 2 μg/L. The rat oral LD₅₀s are 25 and 60 mg/kg. Deltamethrin is the most effective and persistent of the pyrethroids but is highly toxic to mammals.

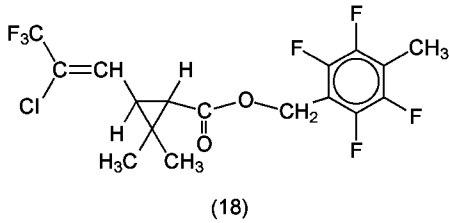
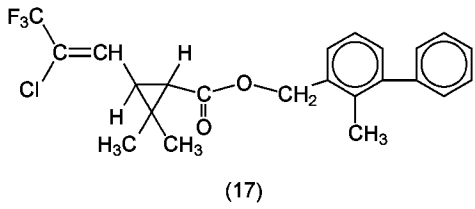


Cyhalothrin [68085-85-8] (*R,S*)-α-cyano-(3-phenoxyphenyl)methyl (±)-*cis*-(1*R,S*) 3-(2-chloro-3,3,3-trifluoropropenyl)-2,2-dimethylcyclopropanecarboxylate (16) (mp 49°C, vp 0.2 μPa at 25°C), a mixture of four isomers, is soluble in water

to 5 µg L. The rat LD₅₀ values are 79, 56 (oral), 632, 696 (dermal) mg kg. It is a persistent pyrethroid insecticide and acaricide (anti-mite) used on cotton and other field crops. Lambda-cyhalothrin [91465-08-6] is a 1:1 mixture of (Z)-(1R,3R)-S and (Z)-(1S,3S)-R esters.

Bifenthrin [82657-04-3], [2-methyl-(1,1'-biphenyl)-3-yl]methyl (Z)-(1R,S)-cis-3-(2-chloro-3,3,3-trifluoropropenyl)-2,2-dimethylcyclopropanecarboxylate (17) (mp 68–70°C, vp 0.034 mPa at 25°C), a mixture of two isomers is soluble in water to 0.1 mg L. The rat oral LD₅₀ is 54 and the rabbit dermal LD₅₀ is 2000 mg kg. Bifenthrin is a broad-spectrum insecticide with acaricidal properties.

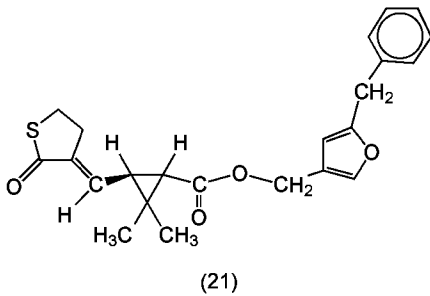
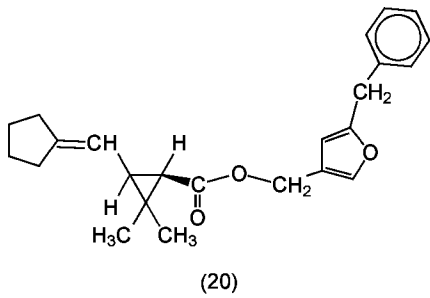
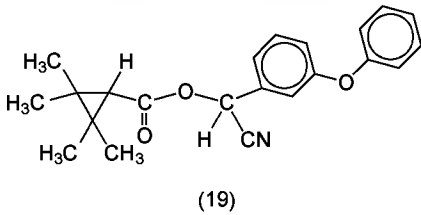
Tefluthrin [79538-32-2], 2,3,5,6-tetrafluoro-4-methylphenylmethyl 2-(±)-(1R,S,3R,S)-cis-(2-chloro-3,3,3-trifluoropropenyl)-2,2-dimethylcyclopropanecarboxylate (18) (mp 44°C, vp 80 mPa at 20°C), a mixture of two isomers, is soluble in water to 20 µg L. The rat LD₅₀s are 23 and 25 (oral) and 148 (dermal) mg kg. Tefluthrin is a persistent soil insecticide.



Fenpropathrin [39515-41-8], α-cyano-(3-phenoxyphenyl)methyl 2,2,3,3-tetramethylcyclopropanecarboxylate (19) (vp 0.73 mPa at 20°C), a mixture of two isomers, is soluble in water to 0.33 mg L. The rat oral LD₅₀ is 25 mg kg and the dermal LD₅₀ >2000 mg kg. Fenpropathrin is an effective acaricide.

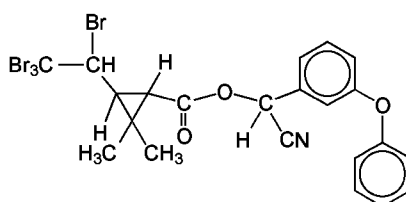
Bioethanomethrin [22431-62-5], 5-(benzyl-3-furanylmethyl)-3-(cyclopentylidenemethyl)-(+)-trans-(1R,3S)-2,2-dimethylcyclopropane carboxylate (20) has a rat oral LD₅₀ of 100 mg kg.

Kadethrin [58769-20-3] is 5-(benzyl-3-furanylmethyl)-3-[dihydro-2-oxo-3(2H)-thienylidene]-methyl]-cis-(1R,3S)-2,2-dimethylcyclopropanecarboxylate (21) (mp 31°C, vp <0.1 mPa at 20°C). The rat LD₅₀s are 140 (oral) and 1300 (dermal). This pyrethroid has a very rapid knockdown and is used to control flying insect pests.

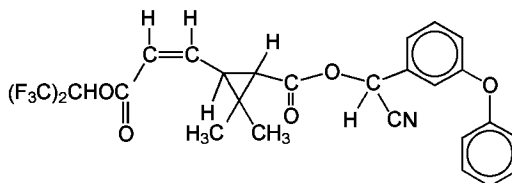


Tralomethrin [66841-25-6], (S)-α-cyano-(3-phenoxyphenyl)methyl (1R,3S)-2,2-dimethyl-3-[(R,S)-1,2,2,2-tetrabromoethyl]cyclopropanecarboxylate (22) (bp 138–148°C, d 1.70, vp 17 mPa at 25°C), a mixture of two isomers, is soluble in water to 70 mg L. The rat oral LD₅₀s are 1250, 1070, and the rabbit dermal LD₅₀ is >2000 mg kg. It is used as a household, stored grain, foliar, and wood protectant insecticide.

Acrinathrin [101007-06-1], (S)-α-cyano-(3-phenoxyphenyl)methyl (Z)-(1R,cis)-2,2-dimethyl-3-{2-[2,2,2-trifluoro-(trifluoromethyl)ethoxycarbonyl]vinyl} cyclopropanecarboxylate (23) (mp 82°C, vp 0.39 µPa at 25°C), is soluble in water to 20 µg L. The rat oral LD₅₀ is >5000 mg kg. It is used for the control of chewing and sucking insects and mites.



(22)

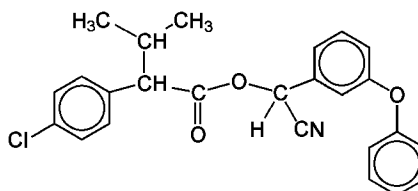


(23)

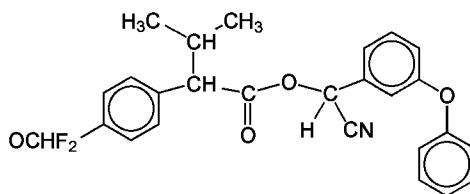
Pyrethroid Esters of Benzene Acetate. These insecticides have more extensive structural optimization in both acid and alcohol moieties. Fenvalerate [51630-58-1], α -cyano-(3-phenoxyphenyl)methyl ($\pm$)-(2*R*,5)- α -isopropyl-4-chlorophenylacetate (**24**) (*d* 1.17, *vp* 1.4 μ Pa at 25°C), a mixture of four isomers, is soluble in water to 0.3 mg/L. The rat oral LD₅₀ is 450 mg/kg. Esfenvalerate [66230-04-4] is the (+)-2-(5*S*)-isomer (mp 59°C). The rat LD₅₀s are 75, 458 (oral), and the rabbit dermal LD₅₀ is 2000 mg/kg. These pyrethroids are widely used general-purpose insecticides for field, vegetable, and fruit crops.

Flucythrinate [70124-77-5], (R,5)- α -cyano-(3-phenoxyphenyl)methyl ($\pm$)-(2*S*)- α -isopropyl-4-difluoromethoxyphenylacetate (**25**) (*d* 1.189, *vp* 1.2 μ Pa at 20°C), a mixture of two isomers, is soluble in water to 0.5 mg/L. The rat oral LD₅₀s are 53, 87 mg/kg, and the rabbit dermal LD₅₀ is >1000 mg/kg.

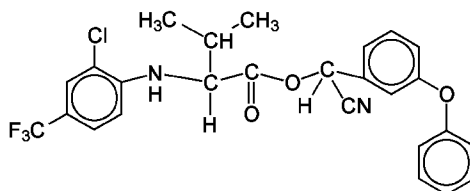
Flucythrinate is a broad-spectrum insecticide.



(24)



(25)



(26)

Fluvalinate [69409-94-5], (R,5)- α -cyano-(3-phenoxyphenyl)methyl *N*-[2-chloro-4-(trifluoromethyl)phenyl] DL-valine (**26**) (*d* 1.29, *vp* <13 μ Pa at 25°C), a mixture of two isomers, is soluble in water to 2 μ g/L. The rat LD₅₀s are >3,000 (oral) and >20,000 (dermal) mg/kg. It is a general-purpose insecticide.

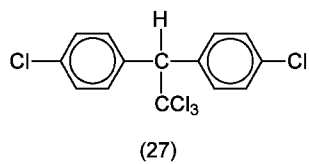
Mode of Action. The pyrethroids readily penetrate the insect cuticle, and their action is characterized by excitation, incoordination, and paralysis, and results in rapid knockdown. The pronounced effects of chirality and cis–trans isomerism upon insecticidal activity and the rigorous structural requirements demonstrate that the natural pyrethroids interact with an axonal receptor site forming a three-point contact at the isobutenyl moiety of the acid, the dimethylcyclopropane ring, and the unsaturated side chain of the keto-alcohol. For the synthetic pyrethroids, the 2-methyl-1-propenyl (isobutenyl) group of the cyclopropanecarboxylic acid moiety or its stereochemical equivalents, eg, the α -isopropylphenylacetic acid, are essential. The C-3 substituent can be varied from isobutenyl to 2,2-dihalovinyl, to trifluoromethylchlorovinyl. The configuration at C-1 and C-3 of the cyclopropanecarboxylic acid must be either 1*R*,3*S* or 1*R*,3*R*. The cyclopentenolone ring of the alcohol moiety of the natural pyrethrins can be structurally optimized to benzyl or furyl groups with an unsaturated side chain or aromatic ether. The α -cyano group appears to be a dimensional spacer to promote maximum complementarity with the axonal receptor. The natural pyrethrins, permethrin, resmethrin, bifenthrin, etc, without the α -cyano group, act at the Na⁺ channels of the nerve axon, causing them to close very slowly following each action potential, thus producing a prolonged after-potential. The pyrethroids with the α -cyano group such as cypermethrin and cyfluthrin cause a delayed closure of the Na⁺ channels to suppress the spontaneous generation of the after-potential. The biochemical lesion is the inhibition of the Ca²⁺-dependent protein kinase associated with axonal Na⁺ channels in the phosphorylation of the neural protein.

Environmental. The development of the synthetic pyrethroids has provided a significant improvement in the nontarget environmental toxicology of insecticides, as these compounds are effective in plant protection at doses of about one-tenth to one-twentieth those of other synthetic insecticides (eg, 4 to 20 g/ha). However, the pyrethroids are extremely toxic to fish, LC₅₀ 1 to 10 μ g/L. They are also very toxic to bees and other beneficial insects. Some of the pyrethroids such as permethrin, cypermethrin, cyfluthrin, and fenvalerate are much more persistent than the natural pyrethrins because degradophores have been removed from the acid and alcohol moieties by the structural optimization process. Thus, whereas the natural pyrethrins and resmethrin have only a one to two day persistence, permethrin, decamethrin, and fenvalerate may persist as foliage residues for two to four weeks and as soil residues for

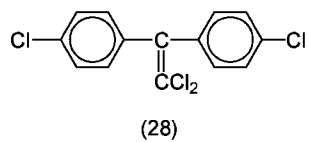
one to two months. The acute toxicity to mammals ranges from very safe insecticides such as the natural pyrethrins, allethrin, resmethrin, tetramethrin, and cyfluthrin, to highly toxic compounds such as deltamethrin, flucythrinate, cyhalothrin, and tefluthrin. These compounds are irritating to the skin and nasal membranes, and the lack of a specific antidote is a disadvantage.

Organochlorine Insecticides. DDT was first synthesized in 1874 but its insecticidal properties were discovered in 1939 (22,23). Its combination of useful properties, ie, low cost, broad-spectrum activity, lengthy persistence, and safety to humans and domestic animals, combined with its immense usefulness in the conquest of insect borne diseases during and immediately after World War II, were responsible for its widespread and enthusiastic use. At the height of its popularity, in 1962, it was registered for use on 334 agriculture commodities in the United States and about 85,000 t were produced. The success of DDT as an insecticide stimulated research and discovery in the organochlorine area. Lindane, toxaphene, chlordane, heptachlor, aldrin, dieldrin, and endrin were developed during the decade after World War II. These organochlorine insecticides dominated the market during the period of 1945–1965 and cumulative world production has been estimated as DDT 2×10^6 t, lindane (as technical BHC) 2.5×10^6 t, toxaphene 4.5×10^6 t, and cyclodienes 2.7×10^6 t (see also CHLOROCARBONS AND CHLOROHYDROCARBONS, TOXIC AROMATICS).

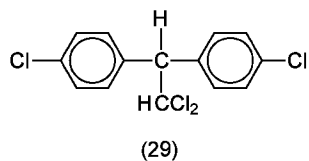
DDT and Analogues. DDT or dichlorodiphenyltrichloroethane is 1,1,1-trichloro-2,2-bis-(4-chlorophenyl)ethane [50-29-3] (27) (mp 109°C, vp 25 µPa at 20°C). It is soluble in water to only 1.2 µg L but is readily soluble in kerosene and fuel oil and in xylene. In alkaline solution, DDT is dehydrochlorinated to form the noninsecticidal DDE, 1,1-dichloro-2,2-bis-(4-chlorophenyl)ethylene [72-55-9] (28) (mp 85°C). This reaction occurs in biological systems and accounts for the predominant storage and bioaccumulation of DDE residues in lipid tissues of humans and animals. DDT is also reductively dechlorinated to form 1,1-dichloro-2,2-bis-(4-chlorophenyl)ethane [72-54-8] (DDD) (29) (mp 110°C) with properties very similar to DDT. Technical DDT (mp 80–94°C) is a white amorphous powder containing 65–80% of the active *p,p'*-isomer, 17–21% of the nearly inactive *o,p'*-isomer [789-02-6] (mp 73°C), and up to 4% *p,p'*-DDD [72-54-8], traces of *o,o'*-DDD and bis-(4-chlorophenyl)sulfone. The rat LD₅₀ values for DDT are 113, 118 (oral), and 2510 (dermal) mg kg. DDD has also been used as an insecticide with rat LD₅₀ values of 3400 (oral) and >10,000 (dermal) mg kg(4,14,22,24).



[50-29-3]



[72-55-9]

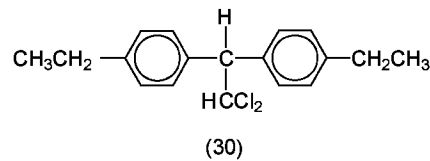


[72-54-8]

Methoxychlor [72-43-5], 1,1,1-trichloro-2,2-bis-(4-methoxyphenyl)ethane, (mp 89°C), where methoxy groups replace the ring chlorines of DDT, is soluble in water to 0.1 mg L. It is much less readily dehydrochlorinated in alkaline solution or in biological systems than is DDT. However, the *p,p'*-methoxy groups are rapidly attacked by microsomal oxidase systems in higher animals to form phenols that are conjugated and eliminated. Thus methoxychlor does not bioaccumulate as does DDT and is favored for general environmental use. Methoxychlor is one of the safest of all insecticides with a rat oral LD₅₀ of >6000 mg kg. It is useful in the home garden, for the control of insect pests of vegetable and fruits, for veterinary hygiene, and for the control of bark beetles that are the vectors of Dutch elm disease.

Perthane [72-56-0], 1,1-dichloro-2,2-bis-(4-ethylphenyl)ethane (30) (mp 60–61°C) is a rapidly biodegradable insecticide with very low mammalian toxicity, rat oral LD₅₀ 8170 mg kg. It has been used as a household insecticide.

The fluorine analogue of DDT, 1,1,1-trichloro-2,2-bis-(4-fluorophenyl)ethane [475-26-3] (DFDT) (mp 44°C), has very similar properties to the chloro compound. It was used as a sanitary insecticide by the German army in World War II, but is too phytotoxic for agricultural use. The rat oral LD₅₀ is 900 mg kg.



Mode of Action. DDT and its analogues specifically affect the peripheral sense organs of insects and produce violent trains of afferent impulses that result in hyperactivity, convulsions, and paralysis. Death results from metabolic exhaustion and the production of an endogenous neurotoxin. The very high lipophilic nature of these compounds facilitates absorption through the insect cuticle and penetration to the nerve tissue. The specific site of action is thought to be the sodium channels of the axon, through inhibition of Ca²⁺ ATPase.

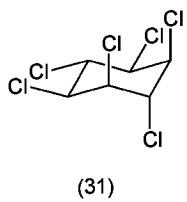
Environmental. DDT is the most permanent and durable of the commonly used contact insecticides because of its insolubility in water, its very low vapor pressure, and its resistance to destruction by light and oxidation. Residues of DDT applied to indoor locations may remain effective for as long as a year, losing activity only when covered by accumulations of grease and dirt. Applications on foliage are less durable, owing to weathering and slow decomposition. The unusual stability of DDT and its high lipid H₂O partitioning (>1 × 10⁶) has resulted in many environmental problems, eg, soil persistence with a half-life of 2.5–10 yr, bioaccumulation from water to fish at levels >1,000,000, transport through food chains, and ubiquitous tissue storage in humans and animals. These properties are shared by DDE, which is even more environmentally recalcitrant.

DDT is slowly converted *in vivo* by reductive dechlorination to DDD and by further dechlorinations to 4,4'-dichlorodiphenylacetic acid [83-05-6] (DDA), the predominant excretory metabolite. Anaerobically, it may form 4,4'-dichlorodiphenylacetonitrile [20968-04-1] (DDCN). However, most DDT that enters the environment is sequestered as DDE, which is ubiquitously present in the body lipids of invertebrate and vertebrate animals. In humans,

DDT is stored in body fats and is secreted in milk as DDT, DDD, and DDE with traces of their *o,p'*-isomers. Levels of these compounds in the body tissues of United States inhabitants have declined slowly from ca 12 ppm in 1970 as a result of sharply curtailed usage.

DDT is highly toxic to fish (LC₅₀ for trout and blue gill, 0.002–0.008 ppm), and it is only moderately toxic to birds (oral LD₅₀ mallard 1300 and pheasant >2240 mg kg). However, widespread bird kills have resulted from bioconcentration of DDT through food chains, ie, from fish or earthworms. A significant environmental problem has resulted from the specific effects of DDE on eggshell formation in raptorial birds where accumulation has caused decreases in shell thickness of 10–15%, resulting in widespread breakage.

Benzene Hexachloride and Lindane. The active constituent of benzene hexachloride is γ -1,2,3,4,5,6-hexachlorocyclohexane [58-89-9] or lindane (31).



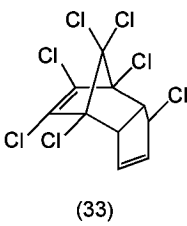
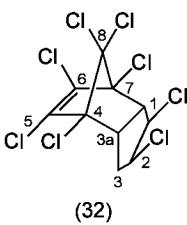
Benzene hexachloride is prepared by the chlorination of benzene in the presence of sunlight. The crude product is a grayish or brownish amorphous solid with a characteristic odor; it begins to melt at 65°C. It consists of 10–18% of the active γ isomer (configuration *aaaaee*, where *a* denotes axial and *e* equatorial) (mp 112°C) with at least four other nearly inactive stereoisomers: α -isomer (*aaaaee* or *aeveea*) (mp 157°C, 55–70%); β isomer (*eeveee*) (mp 309°C, 5–14%); δ isomer (*aeveee*) (mp 138°C, 6–8%); ϵ isomer (*aeveee*) (mp 219°C, 3–4%); and a trace of ν isomer (*aeaaaa*) (mp 90°C) (see CHLOROCARBONS AND CHLOROHYDROCARBONS, CHLORINATED BENZENES). A heptachlorocyclohexane is present up to 4% as a trace of an octachlorocyclohexane; both are insecticidally inactive. The γ isomer is by far the most toxic of the isomers, being from 500–1000 times as active as the α isomer and ca 5000–10,000 times as active as the δ isomer; the β isomer and ϵ isomer are nontoxic. The various isomers differ greatly in their solubilities, and pure γ isomer can be prepared by treating the crude product with methyl alcohol or acetic acid, in which the α and β isomers are nearly insoluble (leaving a marketable product containing 30–40% γ isomer), and then fractionally crystallizing the alcohol-soluble fraction from chloroform or by chromatographic adsorption. The pure γ isomer has a slight aromatic odor (*d* 1.85 g cm⁻³, vp 1.3 mPa at 20°C). It is very stable to the action of heat, light, and oxidation and can be burned without appreciable decomposition but readily decomposed by alkaline materials to form, principally, 1,2,4-trichlorobenzene and three moles of hydrogen chloride. The γ -isomer is soluble in water to 7.3 mg L and is soluble in aromatic solvents. The rat LD₅₀s are 88, 91 (oral), and 900, 1000 (dermal) mg kg.

Lindane is used predominately as a seed dressing and soil insecticide, for the control of ectoparasites of humans and domestic animals, for the control of locusts and grasshoppers, and as a residual spray to control the *Anopheles* vectors of malaria. Because of its relatively high volatility it is useful to control wood-boring insects of timber, fruit trees, and ornamental plants. The mode of action is not well understood but is thought to be competitive blocking of the γ -aminobutyric acid (GABA) transmitter of synaptic nerve transmission.

Cyclodienes. These are polychlorinated cyclic hydrocarbons with endomethylene-bridged structures, prepared by the Diels-Alder diene reaction. The development of these insecticides resulted from the discovery in 1945 of chlordanes, the chlorinated adduct of hexachlorocyclopentadiene and cyclopentadiene (qv). The addition of two Cl atoms across the double bond of the five-membered ring forms the two isomers of chlordanes [12789-03-6] or 1,2,4,5,6,7,8,8-octachloro-2,3,3a,4,7,7a-hexahydro-4,7-methano-1H-indene, α -*trans* (mp 106.5°C) and β -*cis* (32) (mp 104.5°C). The β -isomer has significantly greater insecticidal activity. Technical chlordanes is an amber liquid (bp 175°C 267 Pa, vp 1.3 mPa at 25°C) which is soluble in water to about 9 μ g L. It has rat LD₅₀s of 335, 430 (oral) and 840, 690 (dermal) mg kg. Technical chlordanes contains about 60% of the isomers and 10–20% of heptachlor. It has been used extensively as a soil insecticide for termite control and as a household insecticide.

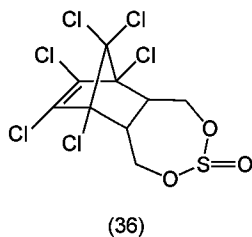
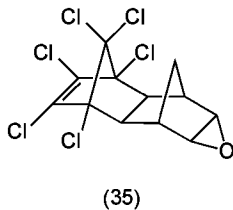
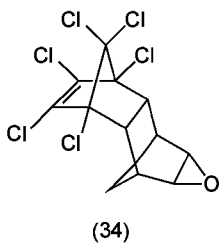
Heptachlor [76-44-8] or 1,4,5,6,7,8,8-heptachloro-3a,4,7,7a-tetrahydro-4,7-methano-1H-indene (33) (mp 95°C, vp 0.04 Pa at 25°C), is soluble in water to 56 μ g L. It is about 3–5 times more active than chlordanes as an insecticide. The rat LD₅₀s are 100, 162 (oral) and 195, 250 (dermal) mg kg. Heptachlor is oxidized readily to heptachlor epoxide [1024-57-3] (mp 159°C), rat oral LD₅₀ 47 mg kg, which is an important and highly persistent environmental pollutant. Hydrogenation of heptachlor produces β -dihydroheptachlor [14168-01-5] (mp 135°C), which retains high insecticidal activity with very low mammalian toxicity, rat oral LD₅₀ >5000 mg kg.

Aldrin [309-00-2] is the adduct of hexachlorocyclopentadiene with bicycloheptadiene, 1,2,3,4,10,10-hexachloro-1,4,4a,5,8,8a-hexahydro-1,4-*endo,exo*-5,8-dimethanonaphthalene (mp 104°C, vp 5.2 mPa at 20°C), water solubility 27 μ g L. The rat LD₅₀s are 39, 60 (oral) and 90 (dermal) mg kg.



Dieldrin [60-57-1] or 1,2,3,4,10,10-hexachloro-1,4,4a,5,8,8a-hexahydro-6,7-epoxy-1,4-*endo,endo*-5,8-dimethanonaphthalene (34) (mp 176°C, vp 0.4 mPa at 20°C) is formed from aldrin by epoxidation with peracetic or perbenzoic acids. It is soluble in water to 27 μ g L. Aldrin and dieldrin have had extensive use as soil insecticides and for seed treatments. Dieldrin, which is very persistent, has had wide use to control migratory locusts, as a residual spray to control the *Anopheles* vectors of malaria, and to control tsetse flies. Because of environmental persistence and propensity for bioaccumulation, registrations in the United States were canceled in 1974.

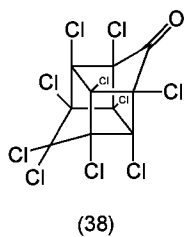
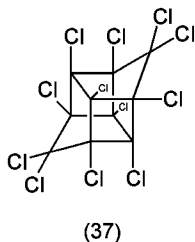
Endrin [72-20-8] is 1,2,3,4,10,10-hexachloro-1,4,4a,5,8,8a-hexahydro-6,7-epoxy-1,4-*endo,endo*-5,8-dimethanonaphthalene (35) (mp 245 dec, vp 0.022 mPa at 25°C) and is soluble in water to 23 μ g L. It is produced by a Diels-Alder reaction of hexachloronorbomadiene with cyclopentadiene, followed by epoxidation. This reaction produces the *endo,endo* isomer of dieldrin, which is less stable and more toxic with rat LD₅₀ values of 17.8 and 7.5 (oral) and 15 (dermal) mg kg. It is used as a cotton insecticide but because of its high toxicity to fish it has been restricted.



Endosulfan [115-29-7] (36) is the adduct of hexachlorocyclopentadiene and 1,4-dihydroxy-2-butene which reacts subsequently with SOCl_2 to produce 6,7,8,9,10,10-hexachloro-1,5,5a,6,9,9a-hexahydro-6,9-methano-2,4,3-benzodioxathiepin-3-oxide. The technical product is a brownish solid (mp 70–100°C, vp 1.3 mPa at 30°C) which consists of about four parts of α -isomer (mp 108°C, cis with regard to the sulfite group) and one part of the β -isomer (mp 206°C, trans with regard to the sulfite group). The α -isomer, which is somewhat more insecticidal, is slowly converted to the more stable β -isomer at high temperature, and both isomers are oxidized slowly to endosulfan sulfate [1031-07-8] (mp 181°C). In acid media, both isomers form endosulfan diol [2157-19-9] (mp 203°C). The rat LD_{50} s are 43, 18 (oral) and 130, 74 (dermal) mg/kg. Endosulfan is a broad-spectrum insecticide for vegetables, fruits, and row crops. Unlike the other cyclodiene insecticides, it is biodegradable by hydrolysis at the sulfite ester bonds.

Mirex [2385-85-5] is 1,2,3,4,5,5,6,7,8,9,10,10-dodecachloro-octahydro-1,3,4-metheno-2H-cyclobuta-[c,d]-pentalene (37) (mp 485°C). The rat LD_{50} s are 306, 600 (oral) and >2000 (dermal) mg/kg. Mirex is extremely resistant to biodegradation and was once considered the perfect stomach poison insecticide for use in baits to control imported fire ants. However, even at doses of a few milligrams per 10 m² it was found to bioaccumulate in birds and fish and its registrations were canceled in the United States in 1976.

Chlordecone [143-50-0] or decachloro-5-oxo-pentacyclo-[5.3.0.0^{2,7}.0^{3,9}.0^{4,8}]-decane (38) (mp 349°C dec) is the 2-keto analogue of mirex and is soluble in water to 4 g/L by hydration. The rat LD_{50} s are 95, 140 (oral) and >2000 (dermal) mg/kg. Chlordecone is a stomach poison used in baits for the control of cockroaches and ants and for the control of banana thrips. Because of bioaccumulation its registrations were canceled in the United States in 1978.



Mode of Action. The cyclodienes, like lindane and toxaphene, affect the nerve axon producing hyperactivity, convulsions, prostration, and death. The biochemical lesion is the competitive inhibition of the γ -aminobutyric acid (GABA) neurotransmitter binding site of the nerve axon. Spray workers with lengthy exposure to dieldrin have suffered from prolonged and repeated central nervous system disturbances producing epileptiform convulsions. Similar disturbances occurred in workers heavily exposed to chlordecone.

Environmental. The high lipophilicity of the cyclodienes and the prolonged persistence of dieldrin and heptachlor epoxide (soil half-lives 2–10 yr) have resulted in severe environmental contamination. These compounds are bioaccumulated from water to fish up to 100,000- to 300,000-fold and are ubiquitous in human fat and milk. Oxychlordan [26880-48-8], mirex, and chlordecone are also bioaccumulative. The cyclodienes are extremely toxic to fish with LC_{50} s (ppm) to trout and bluegill of endrin, 0.001–0.002; endosulfan, 0.001–0.003; dieldrin, 0.003–0.015; aldrin, 0.006–0.01; heptachlor, 0.03–0.026; and chlordan, 0.022–0.095. The LD_{50} s to pheasant and mallard are aldrin 16.8 and 520, dieldrin 79 and 381, and endrin 1.6 and 5.6 mg/kg. As indicated by their rat oral LD_{50} s, they are also extremely toxic to small mammals; in fact, endrin has been used as a rodenticide (see PESTICIDES). Compounds, eg, aldrin and heptachlor, which have unsubstituted double bonds, readily add oxygen to form epoxides in plant and animal tissues and are preferentially concentrated and stored in animal fats. Aldrin epoxide (dieldrin) and heptachlor epoxide are more stable (half-lives on alfalfa of seven to eight days) than aldrin and heptachlor (half-lives on alfalfa of less than one day).

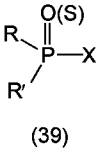
Chlorinated Terpenes. A group of incompletely characterized insecticidal compounds has been produced by the chlorination of the naturally occurring terpenes. Toxaphene [8001-35-2] is prepared by the chlorination of the bicyclic terpene, camphene [79-92-5] to contain 67–69% chlorine and has the empirical formula C₁₀H₁₀Cl₈. The technical product is a yellowish, semicrystalline gum (mp 65–90°C, *d* 1.64) and is a mixture of 175 polychloro derivatives. Toxaphene is unstable in the presence of alkali, upon prolonged exposure to sunlight, and at temperatures above 155°C, liberating hydrogen chloride and losing some of its insecticidal potency. It is very soluble in organic solvents, but only soluble to 0.4 mg L in water. The oral LD₅₀ to the rat is 69 mg kg.

The most active ingredients in technical toxaphene are 2,2,5-*endo*-6-*exo*-8,9,10-heptachlorobornane [51775-36-1] (mouse ip LD₅₀ 6.6 mg kg) and 2,2,5-*endo*-6-*exo*-8,9,9,10-octachlorobornane [58002-18-9] (mouse ip LD₅₀ 3.1 mg kg). Each constitutes ca 2–6% of the technical mixture.

Environmental. Toxaphene is extremely toxic to fish LC₅₀ values to trout and bluegill of 0.003–0.006 ppm. At water concentrations as low as 0.00005 ppm, toxaphene-treated fish suffer broken-back syndrome, a crippling collagen deformity. Bioaccumulation occurs from water to fish at levels up to 100,000-fold. Toxaphene also is highly toxic to birds (oral LD₅₀ to pheasant 40 and 71 mg kg). The soil persistence of toxaphene is difficult to assess because of the complex mixture but published estimates for half-life range from 2 months to 10 years. Toxaphene has been found to be carcinogenic in rats and mice. Toxaphene is a broad-spectrum, persistent pesticide that was widely used on cotton and other field crops. Its registration was revoked by the U.S. Environmental Protection Agency in 1983.

Organophosphorus Insecticides. The discovery of the biological activity of the organophosphorus esters began with the chance observation in 1932 that the vapor of diethylphosphorofluoridate produced cholinergic symptoms in humans. The practical development of this class of compounds as insecticides followed from investigations during World War II, including the development of tetraethyl pyrophosphate as a substitute for nicotine to control aphids (27). This was followed by the lime-stable parathion to control the Colorado potato beetle; schradan, dimefox, and demeton, the first practical plant systemic insecticides; azinphos-methyl, a broad-spectrum insecticide for tree fruits; trichlorfon, a stomach poison; and coumaphos for animal parasites. These successes attracted many other investigators and presently there are more than 100 commercially available organophosphorus (OP) insecticides with a total world production of the order of 2 × 10⁶ t annually (17,25,26,28–30).

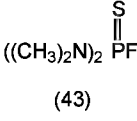
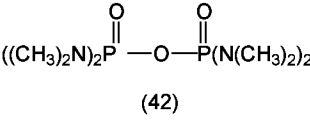
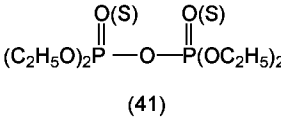
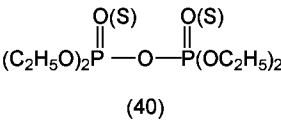
It is estimated that more than 25 × 10⁶ different potentially toxic OP esters can be made using Schrader's classic (27) formula for effective phosphorylating agents, (39), where R and R' are short-chain alkyl, alkoxy, alkylthio, or alkylamino groups, and X is a displaceable moiety with a high energy P-bond such as F or acyl anhydride, and the pentavalent phosphorus atom is bonded to oxygen or sulfur.



Organophosphorus insecticides are available with very short residual action, eg, tetraethyl pyrophosphate and mevinphos, or with prolonged residual activity, eg, diazinon and azinphos-methyl. Parathion, methyl parathion, malathion, and chlorpyrifos are broad-spectrum general-purpose insecticides registered for hundreds of uses. There are OP insecticides with highly specific actions such as dimefox and trichlorfon. The unique properties of demeton and dimethoate have resulted in successful plant systemic insecticides, and this activity has been further refined into compounds such as phorate, terbufos, and disulfoton for seed and soil treatments to protect seedlings. Cruformate and famphur are animal systemic insecticides effective against cattle grubs and other animal parasites. By considering differences in the target enzyme, acetylcholinesterase, and in the different detoxication processes in Mammalia and Insecta, highly selective insecticides such as malathion, fenitrothion, tetrachlorvinphos, and phoxim have been developed. Thus the organophosphorus insecticides represent the most versatile class of chemicals employed for insect control.

Phosphoric Acid and Phosphorothioic Acid Anhydrides. The aliphatic organophosphorus esters originally developed by Schrader (27) are extremely toxic to mammals and are largely of historic interest. Tetraethyl pyrophosphate [107-49-3] (40) (bp 104–110°C at 10.7 Pa, *d* 1.185, vp 6.1 mPa at 30°C) is miscible with water and hydrolyzes very rapidly with a half-life of 6.8 h at 25°C. The rat LD₅₀s are 1.1 (oral) and 2.4 (dermal) mg kg.

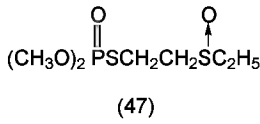
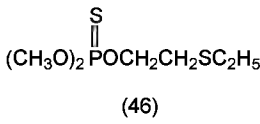
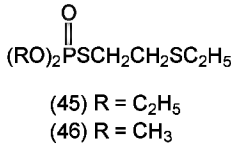
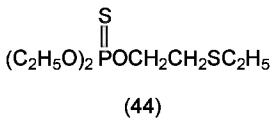
Sulfotepp [3689-24-5], O,O,O',O'-tetraethyl dithiopyrophosphate (41) (bp 110–113°C at 29 Pa, *d* 1.196, vp 22 mPa at 20°C), is soluble in water to 25 mg L. The rat oral LD₅₀ is 5 mg kg. It is used as an aerosol to control greenhouse pests. Schradan [152-16-9], octamethyl pyrophosphoramidate (42) (bp 154°C at 0.5 kPa, vp 0.13 Pa at 25°C), is miscible with water. The rat LD₅₀s are 9.1, 42 (oral) and 15, 44 (dermal) mg kg. Schradan is effective as a systemic insecticide for mites and aphids. Dimefox [115-26-4], N,N,N',N'-tetramethyl phosphorodiamidofluoridate (43) (bp 67°C at 0.5 kPa, *d* 1.12 vp 53 Pa at 30°C), is miscible with water. The rat oral LD₅₀ is 1 mg kg. Dimefox was the first successful systemic insecticide.



Aliphatic Phosphorothioate Esters. Many of the early developments of OP insecticides were simple esters of phosphorothioic acid, (HO)₂P(O)SH, and phosphorodithioic acid, (HO)₂P(S)SH.

Demeton consists of a 2:1 mixture of demeton O [8065-48-3], O,O-diethyl O-(2-ethylthio)ethyl phosphorothionate (44) (thiono or P=S isomer) (bp 123°C at 0.13 kPa, *d* 1.119), water soluble to 66 mg L, rat oral LD₅₀ 30 mg kg; and demeton S [126-75-0], O,O-diethyl S-(2-ethylthio)ethyl phosphorothioate (45) (thiolo or P=O isomer) (bp 128°C at 0.13 kPa, *d* 1.132), soluble in water to 2 g L, rat oral LD₅₀ 1.5 mg kg. The commercial mixture

has rat LD₅₀s of 6.2, 2.5 (oral) and 14, 8.2 (dermal) mg kg. It was the first practical and widely used systemic insecticide for ornamentals and nursery stock.

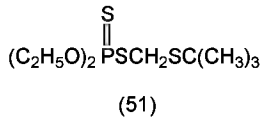
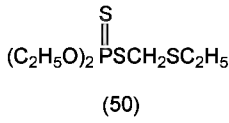
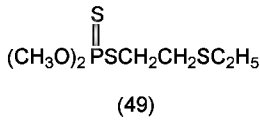
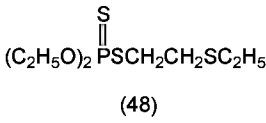


Demeton–methyl is a mixture of *O,O*-dimethyl *O*-(2-ethylthio)ethyl phosphorothioate (demeton–methyl O [8022-00-2] (46)) (bp 74°C at 20 Pa), soluble in water to 330 mg L, and *O,O*-dimethyl *S*-(2-ethylthio)ethyl phosphorothioate (demeton–methyl S) (bp 89°C at 20 Pa), soluble in water to 3.3 g L. The rat oral LD₅₀s are demeton O 180 and demeton S 40, 60 mg kg. Demeton–methyl is a less persistent and safer systemic insecticide.

Oxydemetonmethyl, [301-12-2] *O,O*-dimethyl *S*-(2-ethylsulfinyl)-ethyl phosphorothioate (47), is the sulfinyl derivate of demeton S, rat LD₅₀s 47, 52 (oral) and 173, 158 (dermal) mg kg. It is a rapidly acting systemic insecticide of short persistence.

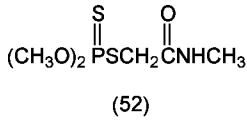
The phosphorodithioic acid esters related to demeton are much more persistent and less water-soluble systemic insecticides. They are widely used as granular products for soil treatments and as dustless seed treatments.

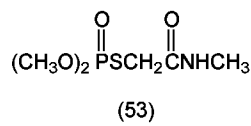
Disulfoton [298-04-4], *O,O*-diethyl *S*-(2-ethylthio)ethyl phosphorodithioate (48) (bp 62°C at 1.3 Pa, *d* 1.144, vp 24 mPa at 20°C), is soluble in water to 66 mg L. The rat LD₅₀s are 6.8, 2.3 (oral) and 25, 6 (dermal) mg kg. Thiometon [144-41-2], *O,O*-dimethyl *S*-(2-ethylthioethyl) phosphorodithioate (49) (bp 110°C at 13 Pa, *d* 1.209, vp 23 mPa at 20°C), is soluble in water to 200 mg L. The rat oral LD₅₀ is 85 mg kg.



Phorate [298-02-2], *O,O*-diethyl *S*-(ethylthio)methyl phosphorodithioate (50) (bp 118–120°C 0.1 kPa, *d* 1.167, vp 0.11 Pa at 20°C), is soluble in water to 85 mg L. The rat LD₅₀s are 2.3, 1.1 (oral) and 6.2, 2.5 (dermal) mg kg. Terbufos [13071-79-9], *O,O*-diethyl *S*-(*t*-butyl)methyl phosphorodithioate (51) (bp 69°C 1.3 Pa, *d* 1.165, vp 35 mPa at 23°C), is soluble in water to about 15 mg L. The rat oral LD₅₀ values are 4.5, 90 mg kg.

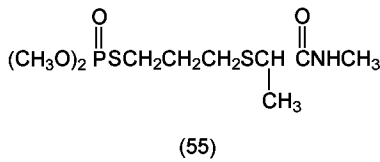
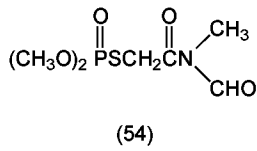
Dimethoate [60-51-5], *O,O*-dimethyl *S*-(*N*-methylcarbamoyl)methyl phosphorodithioate (52) (mp 51°C, *d* 1.277, vp 11 mPa at 25°C), is soluble in water to 70 g L. The rat LD₅₀s are 215, 245 (oral) and 610 (dermal) mg kg. Dimethoate is a widely used contact and systemic insecticide for vegetables, fruits, row crops, and ornamentals. Omethoate [1113-07-6] (53) is the P=O analogue of dimethoate (bp 135°C dec, vp 3.2 mPa at 20°C). The rat LD₅₀ is 50 (oral) and 700 (dermal) mg kg. It is an effective systemic insecticide for control of aphids and red spider mites.





Formothion [2540-82-1], O,O-dimethyl S-(N-formyl-N-methylcarbamoyl)-methyl phosphorodithioate (54) (mp 25°C, *d* 1.36, vp 0.11 mPa at 20°C), is soluble in water to 2.6 g L. The rat LD₅₀s are 365, 500 (oral) and 1000 (dermal) mg kg.

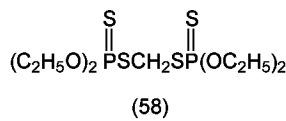
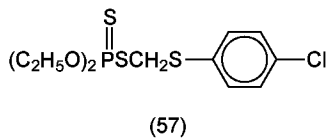
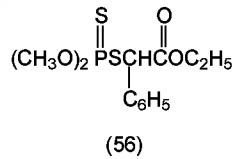
Vamidothion [2275-23-2], O,O-dimethyl-S-[2-(N-methyl-1-methylcarbamoyl)-ethylthio]-ethyl phosphorothioate (55) (mp 46°C), is soluble in water to 40 g L. The rat oral LD₅₀s are 64, 105 mg kg.



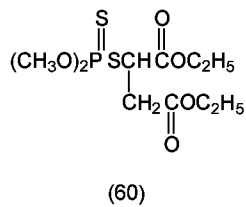
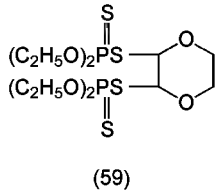
Phenthoate [2597-03-7], O,O-dimethyl S-(α-carboethoxy)benzyl phosphorodithioate (56) (mp 17–18°C, *d* 1.226, vp 5 mPa at 40°C), rat oral LD₅₀s 300, 400 (oral) and 2100 (dermal) mg kg, is a contact and systemic insecticide for tree fruit insects.

Carbophenothion [786-19-6], O,O-diethyl S-(4-chlorophenylthio)methyl phosphorodithioate (57) (bp 82° C 1.3 Pa, *d* 1.271, vp 1.1 mPa at 25°C), is soluble in water to 1 mg L. The rat LD₅₀s are 30, 10 (oral) and 54, 27 (dermal) mg kg. Carbophenothion–methyl is the O,O-dimethyl analogue, rat LD₅₀s 98, 120 (oral) and 215, 190 (dermal) mg kg. These are persistent general-purpose insecticides used on fruit, vegetable, and forage crops.

Ethion [563-12-2], O,O,O',O'-tetraethyl S,S'-methylene diphosphorodithioate (58) (bp 165° C 53 Pa, vp 0.2 mPa at 20°C), has rat LD₅₀s of 65, 27 (oral) and 245, 62 (dermal) mg kg. It is a general-purpose insecticide and acaricide.



Dioxathion [78-34-2], O,O,O',O'-tetraethyl S,S'-(1,4-dioxane-2,3-diyl) diphosphorodithioate (59) (*d* 1.26), has rat LD₅₀s of 43, 23 (oral) and 235, 63 (dermal) mg kg. It is a general-purpose insecticide and acaricide.



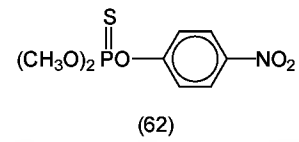
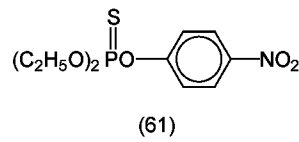
Malathion [121-75-5], O,O-dimethyl S-(1,2-dicarbethoxy)ethyl phosphorodithioate (60) (bp 156–157°C at 93 Pa, *d* 1.23, vp 5.2 mPa at 30°C), is soluble in water to 145 mg L. The rat LD₅₀s are 1375, 1000 (oral) and 4000 (dermal). Malathion readily hydrolyzes in water above pH 7.0 and below pH 5.0. It is one of the most widely used general-purpose insecticides by virtue of its low mammalian toxicity and its good persistence, and is effective for the home garden, household, and against insects of public health importance including flies, mosquitoes, and lice. It is used with protein hydrolysate bait to control fruitflies (Tephritidae).

Phenyl Phosphorothioate Esters. These are the most widely used OP insecticides and incorporate pseudoanhydride high energy phosphate bonds between phosphoric acid and phenols that are present in the activated P=O state.

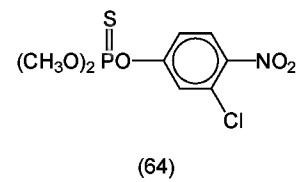
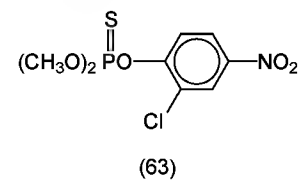
Parathion [56-38-2], O,O-diethyl O-(4-nitrophenyl) phosphorothioate (61) (bp 375°C, *d* 1.265, vp 5.3 mPa at 27°C), is soluble in water to 0.2 mg L.

The rat LD₅₀s are 13, 3.6 (oral) and 21, 6.8 (dermal) mg kg. Parathion is resistant to aqueous hydrolysis, but is hydrolyzed by alkali to form the noninsecticidal diethylphosphorothioic acid and *p*-nitrophenol. The time required for 50% hydrolysis is 120 d in a saturated aqueous solution, or 8 h in a solution of lime water. At temperatures above 130°C, parathion slowly isomerizes to *O,S*-diethyl *O*-(4-nitrophenyl) phosphorothioate [597-88-6] which is much less stable and less effective as an insecticide. Parathion is readily reduced, eg, by *Bacillus subtilis* in polluted water and in the mammalian rumen to nontoxic *O,O*-diethyl *O*-(4-aminophenyl) phosphorothioate, and is oxidized with difficulty to the highly toxic paraoxon [311-45-5], diethyl 4-nitrophenyl phosphate (*d* 1.268, soluble in water to 2.4 mg L), rat oral LD₅₀ 1.2 mg kg.

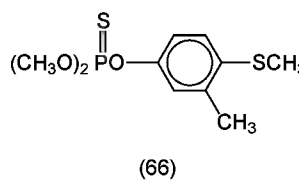
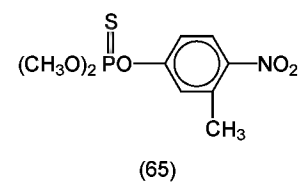
Methyl parathion [298-00-0], *O,O*-dimethyl *O*-(4-nitrophenyl) phosphorothioate (62) (mp 34°C, *d* 1.358, vp 1.3 mPa at 23°C), is soluble in water to 50 mg L. The rat LD₅₀s are 14 and 24 (oral) and 67 (dermal) mg kg. Methyl parathion hydrolyzes and isomerizes more readily than parathion and is favored for use because of its decreased toxic hazard.



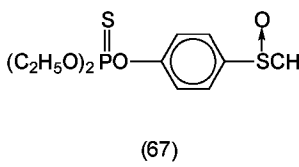
The hazards of human poisoning by the parathions have stimulated the development of safer analogues. Two chlorinated derivatives have greatly reduced mammalian toxicities. Dicapthon [2463-84-5], *O,O*-dimethyl *O*-(2-chloro-4-nitrophenyl) phosphorothioate (63) (mp 53°C), has rat LD₅₀s of 400, 330 (oral) and 790, 1250 (dermal) mg kg. Chlorthion [500-20-8], *O,O*-dimethyl *O*-(3-chloro-4-nitrophenyl) phosphorothioate (64) (mp 21°C, *d* 1.437), has rat LD₅₀s of 890, 980 (oral) and 4500, 4100 (dermal) mg kg. These compounds have been used as household insecticides.

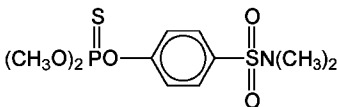


Fenitrothion [122-14-5], *O,O*-dimethyl *O*-(3-methyl-4-nitrophenyl) phosphorothioate (65) (bp 140–145°C at 13 Pa, *d* 1.3227, vp 0.15 mPa at 20°C), is soluble in water to 20 mg L. The rat LD₅₀s are 740, 570 mg kg (oral) and 890, 1200 mg kg (dermal). Fenitrothion is a broad-spectrum insecticide similar to methyl parathion and has been used in residual house spraying for malaria control. Fenthion [55-38-9], *O,O*-dimethyl *O*-(3-methyl-4-methylthiophenyl) phosphorothioate (66) (bp 105°C at 1.3 Pa, *d* 1.256, vp 3.9 mPa at 20°C), is soluble in water to 55 mg L. The rat LD₅₀s are 215, 245 (oral) and 330 (dermal) mg kg. Fenthion is a persistent insecticide used in veterinary hygiene and in mosquito larviciding where it is especially effective in polluted water. It has also been used to control pest birds. Fenthion is activated by microsomal oxidases which convert the CH₃S— group to CH₃SO— and CH₃SO₂—.



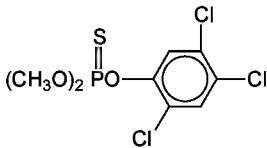
Fensulfothion [115-90-2], *O,O*-diethyl *O*-(4-methylsulfinylphenyl) phosphorothioate (67) (bp 138–141°C at 1.3 Pa), has rat oral LD₅₀s of 4.1, 1.8 (oral) and 19, 4.1 (dermal) mg kg. It is soluble in water to 1.5 g L and is a soil insecticide and nematocide. Famphur [52-85-2], *O,O*-dimethyl *O*-[4-(dimethylsulfamoyl)phenyl] phosphorothioate (68) (mp 53°C) has rat oral LD₅₀s of 35, 65 mg kg and rabbit dermal LD₅₀ 1400 mg kg. It is an animal systemic insecticide.



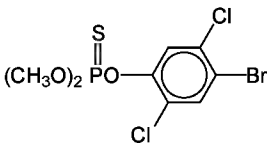


(68)

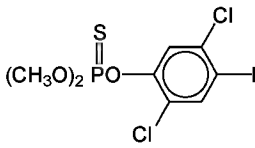
Ronnel [299-84-3] or fenclorophos, *O,O*-dimethyl *O*-(2,4,5-trichlorophenyl) phosphorothionate (69) (mp 41°C, vp 0.11 Pa at 25°C), is soluble in water to 80 mg L. The rat oral LD₅₀s are 1250 and 2630 mg kg. Ronnel is used in veterinary hygiene, especially for the control of cattle grubs. Bromophos [2104-96-3], *O,O*-dimethyl *O*-(2,5-dichloro-4-bromophenyl) phosphorothioate (70) (mp 53°C, vp 17 mPa at 20°C), is soluble in water to 40 mg L. The rat LD₅₀ values are 1600, 1730 (oral) and >5000 (dermal) mg kg. Iodofenphos [18181-70-9], *O,O*-dimethyl *O*-(2,5-dichloro-4-iodophenyl) phosphorothioate (71) (mp 76°C, vp 0.11 mPa at 20°C), is soluble in water to 20 mg L. The rat oral LD₅₀ is 2100 mg kg. These highly selective and relatively safe OP compounds are used in veterinary hygiene and as household insecticides.



(69)



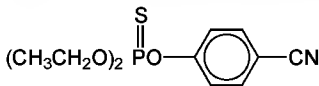
(70)



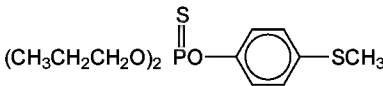
(71)

Cyanophos [2636-26-2], *O,O*-diethyl *O*-(4-cyanophenyl) phosphorothioate (72) (mp 14°C, *d* 1.255, vp 0.5 mPa at 20°C), is soluble in water to 46 mg L. The rat oral LD₅₀s are 580, 610 mg kg and the mouse dermal is 2500 mg kg. Cyanophos is a broad-spectrum, but much safer analogue of parathion.

Propafos [7292-16-2], *O,O*-dipropyl *O*-(4-methylthiophenyl) phosphorothioate (73) (bp 175°C/100 Pa, *d* 1.150, vp 0.12 mPa at 25°C), is soluble in water to 125 mg L. The rat oral LD₅₀ is 70 mg kg. It is used in Japan to control rice leafhoppers resistant to organophosphate and carbamate insecticides.

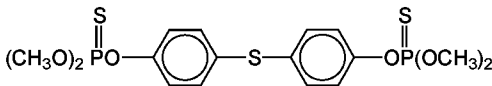


(72)

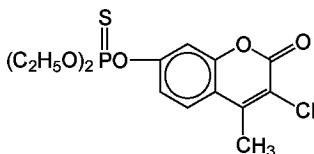


(73)

Temephos [3383-96-8], *O,O,O',O'*-tetramethyl *O,O'*-(thiodi-4,1-phenylene)bisphosphorothioate (74) (mp 30°C, *d* 1.32), is soluble in water to 0.025 mg L. The rat LD₅₀ values are 8,600, 13,000 (oral) and 4,000 (dermal) mg kg. Temephos is a very safe larvicide for mosquitoes and blackflies and has been used for thrips control. Coumaphos [56-72-4], *O,O*-diethyl *O*-(3-chloro-4-methyl-2-oxo-2*H*-1-benzopyran-7-yl) phosphorothioate (75) (mp 91°C, vp 13 μPa at 20°C), has rat LD₅₀s of 41, 16 (oral) and 860 (dermal) mg kg. It is used as a spray or pour-on for control of cattle grubs and ectoparasites.



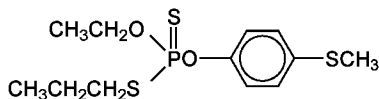
(74)



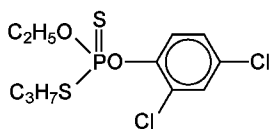
(75)

Phenyl Phosphorodithioate Esters. These differ from the widely used *O,O*-dialkyl esters, by having mixed *O*-alkyl, *S*-alkyl groups.

Sulprofos [35400-43-2], O-ethyl, S-propyl O-[(4-methylthio)phenyl] phosphorodithioate (**76**) (bp 155° C 13 Pa, *d* 1.20, vp 0.1 mPa at 20°C), has rat oral LD₅₀s of 25 and 60 mg kg. It is a persistent broad-spectrum insecticide used for the control of cotton pests. Prothiofos [34643-46-4], O-ethyl, S-propyl O-2,4-dichlorophenyl phosphorodithioate (**77**) (*d* 1.3, vp <1 mPa at 20°C), is soluble in water to 1.7 mg L. It is a persistent, broad-spectrum, nonsystemic insecticide.



(76)



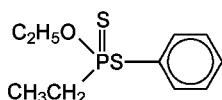
(77)

Phosphonothioate Esters of Phenols. Phosphonates with a single P—C bond are highly toxic and persistent insecticides but have not been used extensively because some compounds produce delayed neuropathy leading to irreversible paralysis in higher animals, including humans. Such compounds specifically inhibit an enzyme, neurotoxic esterase, that is responsible for the growth and maintenance of long nerve axons (31,32).

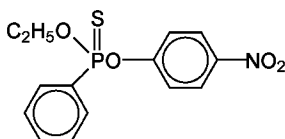
Fonofos [944-22-9], O-ethyl, S-phenyl ethylphosphonodithioate (**78**) (bp 130°C at 13.3 Pa, *d* 1.16 vp 28 mPa at 25°C), is soluble in water to 13 mg L. The rat oral LD₅₀ values are 8 and 17 mg kg. Fonofos is a persistent soil insecticide.

EPN [2104-64-5], O-ethyl O-(4-nitrophenyl) phenylphosphonothioate (**79**) (mp 41°C, *d* 1.268, vp 126 μPa at 25°C), is soluble in water to about 0.1 mg L. The rat LD₅₀s are 36, 7.7 (oral) and 230, 25 (dermal) mg kg. It has been used for the control of cotton pests but is a delayed neurotoxin.

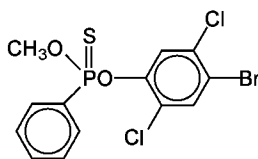
Leptophos [21609-90-5], O-methyl O-(4-bromo-2,5-dichlorophenyl) phenylphosphonothioate (**80**) (mp 71°C), is soluble in water to 0.03 mg L. The rat oral LD₅₀s are 53, 45 mg kg. Leptophos has been extensively used as a cotton insecticide but is a delayed neurotoxin.



(78)



(79)



(80)

Vinyl Phosphates. Dichlorvos [62-73-7], O,O-dimethyl O-(2,2-dichlorovinyl) phosphate, (CH₃O)₂P(O)OCH=CCl₂ (bp 140°C at 27 kPa, *d* 1.314, vp 1.6 Pa at 20°C), is soluble in water to about 10 g L. The half-life in water is 8 h. The rat oral LD₅₀s are 80, 56 mg kg. Dichlorvos is used in aerosols and sugar baits to control flies and mosquitoes. Slow release formulations have been used in plastic strips and pet collars to control animal ectoparasites.

Naled [300-76-6], O,O-dimethyl O-(1,2-dibromo-2,2-dichloroethyl) phosphate, (CH₃O)₂P(O)OCHBrCBrCl₂ (mp 27°C, bp 110°C at 66.7 Pa, *d* 1.96, vp 0.26 Pa at 20°C), is very slightly soluble in water. The rat LD₅₀s are 250 (oral) and 800 (dermal) mg kg. Naled is a rapid acting, nonpersistent insecticide used in sugar baits for flies and protein baits for fruitflies.

Trichlorfon [52-68-6], O,O-dimethyl 1-hydroxy-2,2,2-trichloroethyl phosphonate, (CH₃O)₂P(O)CH(OH)CCl₃ (mp 83°C, *d* 1.73, vp 1 mPa at 20°C), is soluble in water to 154 g L. The rat LD₅₀s are 630, 560 (oral) and >2000 (dermal) mg kg. Above pH 6, trichlorfon rearranges to form dichlorvos, accounting for its toxic action. It is used on foliage as a nonpersistent stomach poison for chewing insects and in dry sugar bait for fly control.

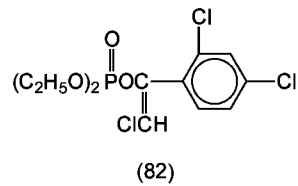
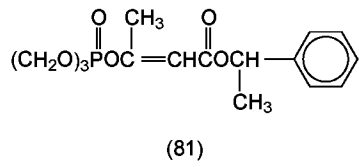
Mevinphos [7786-34-7], O,O-dimethyl O-(2-methoxycarbonyl-1-methylvinyl) phosphate, (CH₃O)₂P(O)C(CH₃)=CHC(O)OCH₃ (bp 106–107°C at 0.13 kPa, *d* 1.25, vp 0.38 Pa at 21°C), is a mixture of one part *cis* and two parts *trans* isomers, of which the *cis* isomer is about 10-fold more effective. It is miscible with water and rat LD₅₀s are 6.1, 3.7 (oral) and 4.7, 4.2 (dermal) mg kg. Mevinphos is a nonpersistent systemic insecticide suitable for the treatment of edible produce close to harvest, because of rapid dissipation of the residue by hydrolysis and volatilization.

Phosphamidon [13171-21-6], O,O-dimethyl O-(2-chloro-2-diethylcarbamoyl-1-methylvinyl) phosphate, (CH₃O)₂P(O)OC(CH₃)=C(Cl)C(O)N(C₂H₅)₂ (bp 160°C at 0.2 kPa, *d* 1.21, vp 3.2 mPa at 20°C), is miscible with water, and the rat LD₅₀s are 24 (oral) and 143, 107 (dermal) mg kg. It is a nonpersistent systemic insecticide and acaricide for tree fruits.

Dicrotophos [141-66-2], *cis*-O,O-dimethyl O-(2-dimethylcarbamoyl)-1-methylvinyl phosphate, (CH₃O)₂P(O)C(CH₃)=CHC(O)N(CH₃)₂ (bp 400°C, *d* 1.216, vp 9.3 mPa at 20°C), is miscible with water. The rat LD₅₀s are 21, 16 (oral) and 43, 42 (dermal) mg kg. It is a contact and systemic insecticide which has been injected into trees to control bark beetles and foliage pests.

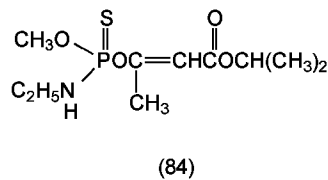
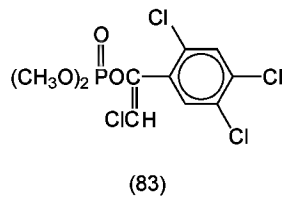
Monocrotophos [6923-22-4], *cis*-*O,O*-dimethyl *O*-(2-methylcarbamoyl)-1-methylvinyl phosphate, (CH₃O)₂P(O)C(CH₃)=CHC(O)NHCH₃ (mp 54°C, *d* 1.33, vp 0.9 mPa at 20°C), is miscible with water. The rat LD₅₀s are 17, 20 (oral) and 126, 112 (dermal) mg kg. Monocrotophos, a metabolite of dicrotophos, is a contact and systemic insecticide.

Crotoxyphos [7700-17-6], *O,O*-dimethyl-*O*-(1-methyl-2-(1-phenylcarbethoxy)vinyl phosphate (**81**) (bp 135°C at 4 Pa, *d* 1.19, vp 1.8 mPa at 20°C), is soluble in water to 1 g L. The rat LD₅₀s are 110, 74 (oral) and 375, 202 (dermal) mg kg. It is used to control animal ectoparasites. Chlorfenvinphos [470-90-6], *O,O*-diethyl-*O*-[2-chloro-1-(2,4-dichlorophenyl)-vinyl] phosphate (**82**) (bp 168–170°C at 67 Pa, *d* 1.36, vp 0.97 mPa at 25°C), is soluble in water to 145 mg L. The rat LD₅₀s are 15, 13 (oral) and 31, 30 (dermal) mg kg. It is a contact and soil insecticide for agricultural pests.



Tetrachlorvinphos or stirifos [22248-79-9], *O,O*-dimethyl *O*-[2-chloro-1-(2,4,5-trichlorophenyl)vinyl] phosphate (**83**) (mp 97–98°C, vp 5 μPa at 20°C), is soluble in water to 11 mg L. The rat LD₅₀s are 1100, 1125 (oral) and >5000 (dermal) mg kg. It is used to protect stored products for veterinary hygiene, and to control vegetable and fruit insects.

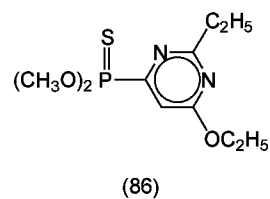
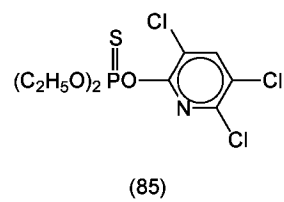
Propetamphos [31218-83-4], *O*-[2-(isopropoxycarbonyl)-1-methylvinyl]-*O*-methyl, *N*-ethyl phosphoramidothioate (**84**) (bp 87–89°C at 0.7 Pa, *d* 1.129, vp 1.9 mPa at 20°C), is soluble in water to 110 mg L. The rat LD₅₀s are 75, 82 (oral) and 2300 (dermal) mg kg. It is used as a household insecticide and to control cotton pests.

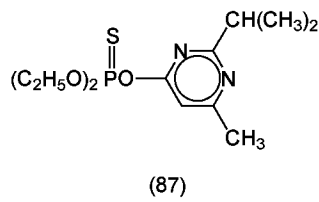


Phosphorothioate Esters of Heterocyclic Enols. Chlorpyrifos [2971-88-2], *O,O*-diethyl *O*-(3,5,6-trichloro-2-pyridinyl) phosphorothioate (**85**) (mp 42–43°C, vp 2.5 mPa at 25°C), is soluble in water to 2 mg L. The rat LD₅₀s are 135, 165 (oral) and 2000 (dermal) mg kg. Chlorpyrifos is a broad-spectrum general-purpose insecticide and is also used as a soil insecticide and for termite control. Chlorpyrifos–methyl is the *O,O*-dimethyl analogue of chlorpyrifos (mp 45–46°C, vp 5.6 mPa at 25°C). It is soluble in water to 5 mg L. The rat LD₅₀ values are 941, 2140 mg kg. It is used as a larvicide for mosquitoes and black flies and to protect stored grains. Fospirate [5598-52-7], dimethyl 3,5,6-trichloro-2-pyridinyl phosphate, is the P=O analogue of chlorpyrifos–methyl (mp 87°C). The rat oral LD₅₀ is 869 mg kg. It is used as an insecticide and nematocide in veterinary hygiene.

Etrimfos [38260-54-7], *O,O*-dimethyl *O*-(6-ethoxy-2-ethyl-4-pyrimidinyl) phosphorothioate (**86**) (*d* 1.195, vp 8.4 mPa at 20°C), is soluble in water to 40 mg L. The rat LD₅₀ values are 1600, 1800 (oral) and 2000 (dermal) mg kg. It is used as a tree fruit insecticide and acaricide.

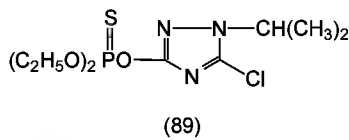
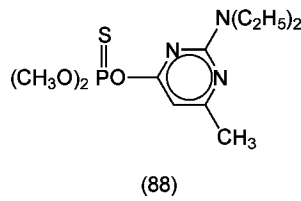
Diazinon [333-41-5], *O,O*-diethyl *O*-(2-isopropyl-4-methyl)-6-pyrimidinyl thiophosphate (**87**) (bp 83–84°C at 0.3 Pa, *d* 1.116, vp 18.2 mPa at 20°C), is soluble in water to 40 mg L. The rat LD₅₀s are 250, 285 (oral) and >2000 (dermal) mg kg. Diazinon is a broad-spectrum insecticide and soil insecticide.



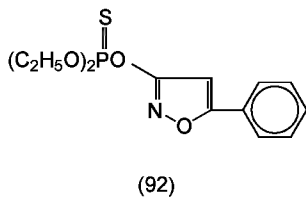
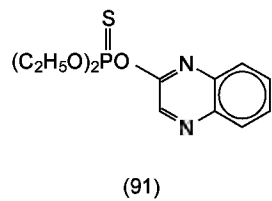
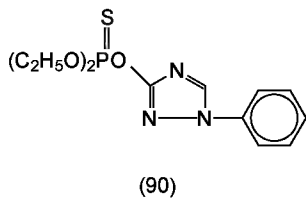


Pirimiphos–methyl or pyrimithate [29232-93-7], *O,O*-dimethyl *O*-(2-diethylamino-6-methyl-4-pyrimidinyl)phosphorothioate (**88**) (*d* 1.157, vp 15 mPa at 30°C), is soluble in water to 5 mg L. The rat LD₅₀ values are 2050 (oral) and 2000 (dermal) mg kg. It is used to control stored product insects and for veterinary hygiene. Pirimiphos–ethyl is the *O,O*-diethyl analogue (bp 128–132°C at 5.3 mPa, *d* 1.165, vp 39 mPa at 25°C) and is soluble in water to 1 mg L. The rat LD₅₀s are 140, 200 (oral) and 1000–2000 (dermal) mg kg. It is an agricultural insecticide and acaricide.

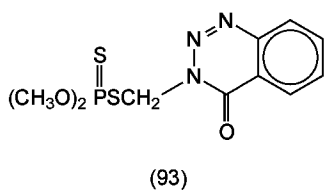
Isazophos [42509-80-8], *O,O*-diethyl *O*-(5-chloro-1-isopropyl-(1*H*)-1,2,4-triazol-3-yl)phosphorothioate (**89**) (bp 100°C at 1.3 Pa, *d* 1.22, vp 4.3 mPa at 20°C), is soluble in water to 150 mg L. The rat LD₅₀s are 40, 60 (oral) and 3100 (dermal) mg kg. It is a soil insecticide and nematocide.

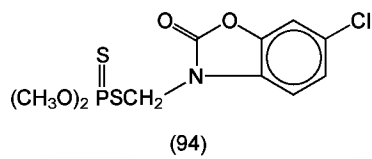


Triazophos [24017-47-8], *O,O*-diethyl *O*-(1-phenyl-1*H*-1,2,4-triazol-3-yl)phosphorothioate (**90**) (*d* 1.247, vp 0.39 mPa at 30°C), is soluble in water to 30 mg L. The rat LD₅₀s are 57, 68 (oral) and >1100 (dermal) mg kg. It is a broad-spectrum insecticide and nematocide. Quinalphos [13593-03-8], *O,O*-diethyl *O*-(quinoxalin-2-yl)phosphorothioate (**91**) (*d* 1.235, vp 0.34 mPa at 20°C), is soluble in water to 2 mg L. The rat LD₅₀s are 70 (oral) and >1750 (dermal) mg kg. It is a broad-spectrum acaricide and insecticide. Isoxathion [18854-01-8], *O,O*-diethyl *O*-(5-phenylisoxazol-3-yl)phosphorothioate (**92**) (bp 180°C at 20 Pa, vp 0.133 mPa at 25°C), is soluble in water to 1.9 mg L. The rat LD₅₀s are 112 (oral) and >450 (dermal) mg kg. It is a general-purpose insecticide.

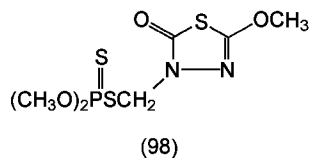
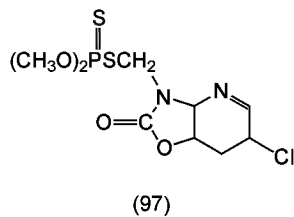
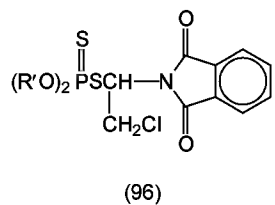
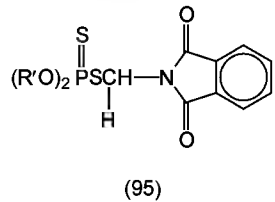


Phosphorothioate Esters of *S*-Methyl Heterocycles. Azinphos-methyl [86-50-0], *O,O*-dimethyl *S*-4-oxobenzo-[d]-(1,2,3)-triazin-3-yl-methylphosphorodithioate (**93**) (mp 73–74°C, *d* 1.44, vp <1 mPa at 20°C), is soluble in water to 30 mg L. The rat LD₅₀s are 13, 11 (oral) and 220 (dermal) mg kg. It is a broad-spectrum persistent insecticide especially used for deciduous fruits. Azinphos–ethyl [2642-71-9] is the corresponding *O,O*-diethyl ester (mp 53°C, *d* 1.284, vp 0.029 mPa at 20°C). It is soluble in water to 6.4 mg L. The rat LD₅₀s are 17.5 (oral) and 250 (dermal) mg kg. Phosalone [2310-17-0], *O,O*-diethyl *S*-(6-chloro-2,3-dihydro-2-oxobenzoxazol-3-yl)-methylphosphorodithioate (**94**) (mp 48°C, vp 0.067 mPa at 25°C), is soluble in water to 10 mg L. The rat oral LD₅₀s are 127, 175 mg kg. It is an insecticide and acaricide for tree fruits.



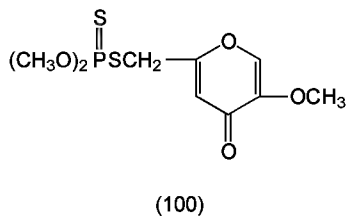
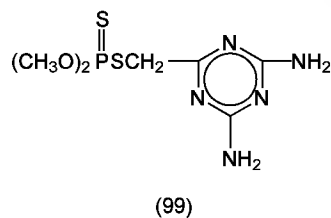


Phosmet [732-11-6], *O,O*-dimethyl *S*-(*N*-phthalimidomethyl) phosphorodithioate (**95**) ($R' = CH_3$) (mp 72°C, vp 0.13 Pa at 50°C), is soluble in water to 24 mg/L. The rat LD₅₀s are 113, 160 (oral) and 1500 (dermal) mg/kg. It is an insecticide for deciduous fruits. Dialifor [10311-84-9], *O,O*-diethyl *S*-(2-chloro-1-phthalimidoethyl) phosphorodithioate (**96**) ($R' = C_2H_5$) (mp 67–69°C, vp 133 mPa at 35°C), is soluble in water to 1 mg/L. The rat LD₅₀s are 43, 75 (oral) and 1500 (dermal) mg/kg. It is an insecticide for deciduous fruits.

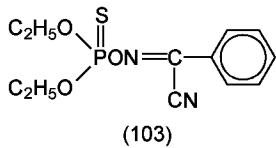
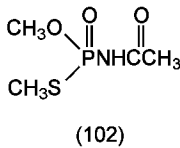
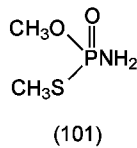


Azamethiphos [35575-96-3], *O,O*-dimethyl *S*-[(6-chloro-2-oxooxazole-[4,5-*b*]-pyridin-3-(2*H*)-3-yl)-methyl] phosphorodithioate (**97**) (mp 89–90°C, vp 5 μPa at 20°C), is soluble in water to 1.1 g/L. The rat oral LD₅₀s are 1100, 1750 mg/kg. It is used for household pest control. Methidathion [950-37-8], *O,O*-dimethyl *S*-[(5-methoxy-2-oxo-1,3,4-thiadiazol-3-(2*H*)-yl)-methyl] phosphorodithioate (**98**) (mp 39–40°C, vp 0.19 Pa at 20°C), is soluble in water to 240 mg/L. The rat oral LD₅₀s are 25, 48 mg/kg. It is a general-purpose insecticide for fruits and vegetables.

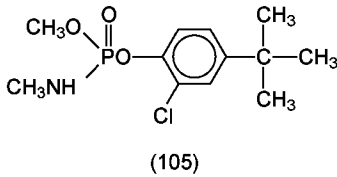
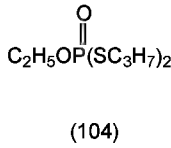
Menazon [78-57-9], *O,O*-dimethyl *S*-(4,6-diamino-1,3,5-triazin-2-yl)-methyl phosphorodithioate (**99**) (mp 160–162°C, vp 0.13 mPa at 25°C), is soluble in water to 2.4 g/L. The rat oral LD₅₀ is 1950 mg/kg. It is a systemic aphicide. Endothion [2778-04-3], *O,O*-dimethyl *S*-[(5-methoxy-4-oxo-4*H*-pyran-2-yl)-methyl] phosphorothioate (**100**) mp 96°C, has rat oral LD₅₀s of 30, 50 mg/kg. It is a systemic insecticide.



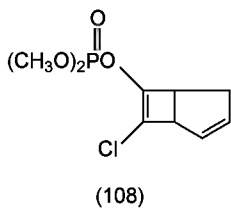
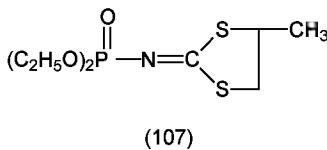
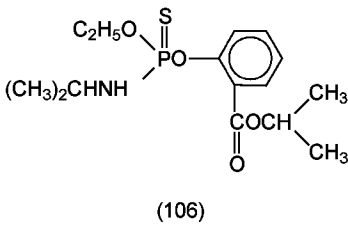
Miscellaneous Organophosphorus Esters. Methamidophos [10265-92-6], *O,S*-dimethylphosphoramidothioate (**101**) (mp 39–41°C, *d* 1.3, vp 39 mPa at 30°C), is soluble in water to about 1 g/L. The rat LD₅₀ values are 19, 21 (oral) and 130 (dermal) mg/kg. It is a contact and systemic insecticide for insect pests of cotton, potato, and vegetable crops. Acephate [30560-19-1], *O,S*-dimethyl *N*-acetyl phosphoramidothioate (**102**) (mp 72–80°C, vp 0.23 mPa at 24°C), is the *N*-acetyl derivative of methamidophos. It is soluble in water to 650 g/L. The rat oral LD₅₀s are 866, 945 mg/kg. Acephate has a much lower toxicity to mammals than methamidophos and is effective against a wide variety of agricultural and household insect pests. Phoxim [14816-18-3], *O,O*-diethyl *O*-(α -cyanobenzilideneamino) phosphorothioate (**103**) (mp 5°C, bp 102°C at 1.3 Pa, *d* 1.176, vp 16 mPa at 20°C), has a rat oral LD₅₀ of 2500 mg/kg. It is used as a grain protectant.



Ethoprophos or ethoprop [13194-48-4], *O*-ethyl, *S,S*-dipropyl phosphorothioate (**104**) (bp 86–91°C at 27 Pa, *d* 1.094, vp 45 mPa at 26°C), is soluble in water to 750 mg L. The rat LD₅₀ values are 61 (oral) and 25 (dermal) mg kg. It is used to control insect and nematode pests of vegetables and ornamentals. Cruformate [299-86-5], *O*-methyl *O*-[(2-chloro-4-(*i*-butyl)phenyl] *N*-methylphosphoramidate (**105**) (mp 61°C), rat oral LD₅₀s 635, 460 mg kg, is an animal systemic insecticide used for cattle grub control. It can be applied either as a 0.75% spray or fed at 20–25 mg kg body weight.



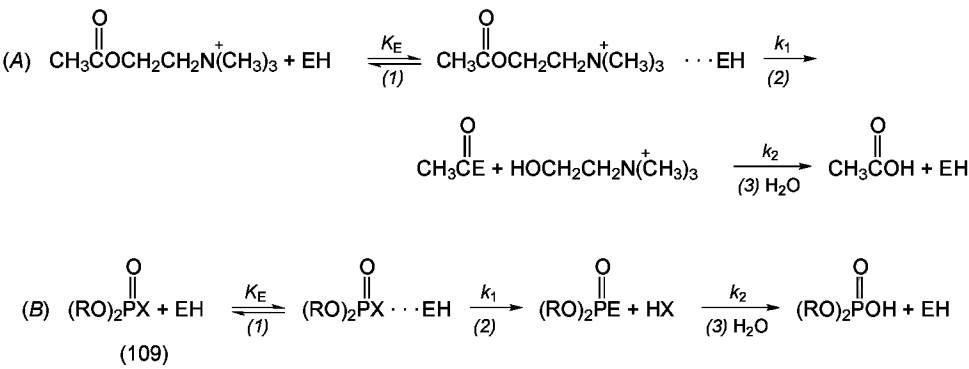
Isofenphos [24311-71-1], *O*-ethyl, *O*-[2-(isopropoxycarbonyl)phenyl]-*N*-isopropyl phosphoramidothioate (**106**) (bp 120°C at 1.3 Pa, *d* 1.13, vp 0.52 mPa at 20°C), is soluble in water to 24 mg L. The rat LD₅₀s are 28 (oral) and 1000 (dermal) mg kg. It is a persistent soil insecticide. Mephosfolan [950-10-7], *O,O*-diethyl *N*-phosphinylimino-4-methyl-1,3-dithiolane (**107**) (bp 120°C at 0.13 Pa), is soluble in water to 57 g L. The rat oral LD₅₀ is 11 mg kg. It has been used as a cotton insecticide and acaricide.



Heptenophos [23560-59-0], *O,O*-dimethyl *O*-(7-chlorobicyclo-[3.2.0]-hepta-2,6-dien-6-yl)phosphate (**108**) (bp 94°C at 0.13 Pa, vp 97 mPa at 20°C), is soluble in water to 2.2 g L. The rat LD₅₀s are 121, 96 (oral) and 2925 (dermal). It is a systemic insecticide with short persistence.

Mode of Action. The organophosphorus insecticides owe their biological activities to the capacity of the central P atom to phosphorylate the esteratic site of the enzyme acetylcholinesterase (AChE), which is an essential constituent of the nervous system of insects as well as higher animals (18,25,26,31–34). The phosphorylated enzyme is irreversibly inhibited and is therefore no longer able to carry out its normal function of the rapid removal and destruction of the neurohormone, acetylcholine (ACh), from the nerve synapse. As a result, ACh accumulates and disrupts the normal functioning of the nervous system, giving rise to the typical cholinergic symptoms associated in insects with OP poisoning, ie, hyperactivity, tremors, convulsions, paralysis, and death. In higher animals, these cholinergic effects are translated into muscarinic effects, eg, nausea, salivation, lacrimation, and myosis; nicotinic effects, eg, muscular fasciculations; and central nervous system effects, eg, giddiness, tremulousness, convulsions, and coma (see ENZYME

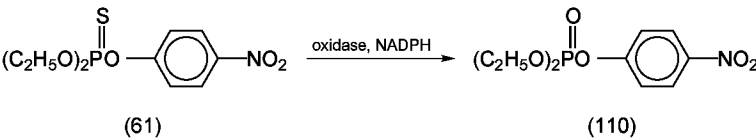
INHIBITORS).



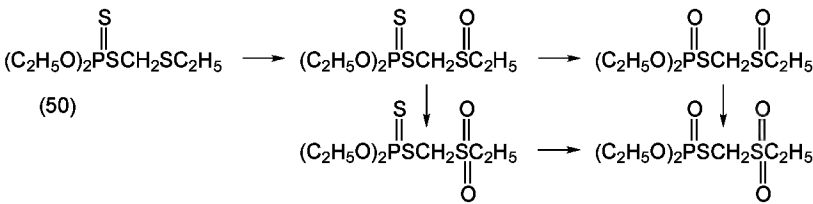
The reaction between esterase and phosphorus inhibitor (109) is bimolecular, of the well-known S_N2 type, and represents the attack of a nucleophilic serine hydroxyl with a neighboring imidazole ring of a histidine residue at the active site, on the electrophilic phosphorus atom, and mimics the normal three-step reaction that takes place between enzyme and substrate (reaction A).

In the normal process (A), step (3) occurs very rapidly and step (1) is the rate-determining step, whereas in the inhibition process (B), step (3) occurs very slowly, generally over a matter of days, so that it is rate determining. Thus it has been demonstrated with AChE that insecticides, eg, tetraethyl pyrophosphate and mevinphos, engage in first-order reactions with the enzyme; the inhibited enzyme is a relatively stable phosphorylated compound containing one mole of phosphorus per mole of enzyme; and as a result of the reaction, an equimolar quantity of alcoholic or acidic product HX is liberated.

The reactivity of the individual O—P insecticides is determined by the magnitude of the electrophilic character of the phosphorus atom, the strength of the bond P—X, and the steric effects of the substituents. The electrophilic nature of the central P atom is determined by the relative positions of the shared electron pairs, between atoms bonded to phosphorus, and is a function of the relative electronegativities of the two atoms in each bond (P, 2.1; O, 3.5; S, 2.5; N, 3.0; and C, 2.5). Therefore, it is clear that in phosphate esters (P=O) the phosphorus is much more electrophilic and these are more reactive than phosphorothioate esters (P=S). The latter generally are so stable as to be relatively unreactive with AChE. They owe their biological activity to *in vivo* oxidation by a microsomal oxidase, a reaction that takes place in insect gut and fat body tissues and in the mammalian liver. A typical example is the oxidation of parathion (61) to paraoxon [311-45-5] (110).



Another type of *in vivo* microsomal oxidation is of importance in the lethal synthesis through which electron-donating groups, eg, CH₃S—, are converted to the strongly electron-withdrawing groups, CH₃SO— and CH₃SO₂— (note the equivalence in toxicity between these three groups in Table 4). This type of reaction (with its built-in delay principle) is involved in the mode of action of certain systemic insecticides, eg, phorate (50) and disulfoton (48), where both P=S and C₂H₅S— oxidation occur.



In this type of activation, which occurs in both animal and plant tissues, the original insecticide is relatively stable and can be translocated through plant tissues without destructive hydrolysis until the oxidation has occurred, which then makes the insecticide both highly toxic and relatively unstable so that it rapidly is hydrolyzed to nontoxic products.

Table 4. Relative Toxicity of Substituted Phenyl Diethyl Phosphates

X	CAS Registry Number	K _E ^a (fly AChE)	Topical LD ₅₀ ^b μg/g	σ ^c
	[311-45-5]	2.7 × 10 ⁷	1.0	1.267
	[6132-17-8]	1.9 × 10 ⁶	2.5	1.049
	[6132-16-7]	2.6 × 10 ⁵	3.5	0.891
	[6552-21-2]	2.4 × 10 ⁵	1.5	0.730
	[5076-63-1]	2.3 × 10 ³	150	0.227

	[2510-86-3]	2.6 × 10	1500	0.0
	[3070-13-1]	2.3 × 10 ³	2.0	0.047
	[4877-08-1]	2.0 × 10	>5000	0.170
	[5076-68-6]	2.9 × 10	>5000	0.135

^a See reactions (A) and (B).

^b *Musca domestica*.

^c Hammett sigma constant (33).

The substituent groups on the phosphorus atom also strongly affect its electrophilic nature, mainly by inductive and mesomeric effects, which may be transmitted along chains of atoms and are positive (electrons repelled) or negative (electrons attracted). Studies with diethyl substituted-phenyl phosphates have shown a high degree of correlation between the quantitative measure of electron-withdrawing power of the substituent group, Hammett's σ, and anticholinesterase activity and toxicity (33) (see Table 4).

The alkyl and alkoxy substituents of phosphate or phosphonate esters also affect the phosphorylating ability of the compound through steric and inductive effects. A satisfactory correlation has been developed between the quantitative measure of these effects, Tafts's σ*, and anticholinesterase activity as well as toxicity (33). Thus long-chain and highly branched alkyl and alkoxy groups attached to phosphorus promote high stability and low biological activity.

Selective Toxicity of Organophosphorus Insecticides. Selectivity is implicit in the definition of insecticide as an agent for destroying insects. Nevertheless, many widely used products, eg, hydrogen cyanide and parathion, are highly toxic to nearly all animals and selectivity is apparent only through careful application and confinement to the area of treatment. The organophosphorus compounds, with their tremendous diversity of structures, provide opportunities for the development of true physiological selectivity. Such selectivity is important not only in providing safety for the user of pesticides and for plant and animal possessions but in providing ecological selectivity for wildlife and for the valuable beneficial insects, ie, honeybees, pollinators, parasites, and predators.

The development of malathion in 1950 was an important milestone in the emergence of selective insecticides. Malathion is from one-half to one-twentieth as toxic to insects as parathion but is only about one two-hundredths as toxic to mammals. Its worldwide usage in quantities of thousands of metric tons in the home, garden, field, orchard, woodland, on animals, and in public health programs has demonstrated substantial safety coupled with pest control effectiveness. The biochemical basis for the selectivity of malathion is its rapid detoxication in the mammalian liver, but not in the insect, through the attack of carboxyesterase enzymes on the aliphatic ester moieties of the molecule.

Extraordinary selectivity has been accomplished with the parathion-type of insecticide by incorporating Cl or CH₃ groups in the meta position of the aryl ring. These groups interact sterically with acetylcholinesterase (AChE), increasing the affinity for the insect enzyme and decreasing it with the mammalian enzyme. Selectivity also is enhanced by differences in the rates of microsomal oxidation of P=S to P=O and in hydrolytic detoxication between insects and mammals. The compounds, fenitrothion, fenthion, ronnel, bromophos, iodofenfos, chlorpyrifos-methyl, and pirimiphos-methyl, and dicapthion are used widely in public health, household, and agricultural pest control.

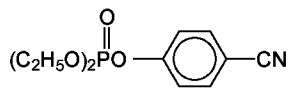
Environmental. Organophosphorus insecticides are intrinsically reactive and readily degrade by oxidation and hydrolysis and in living organisms. Therefore, their use does not present serious problems of biomagnification and food-chain transfer. Soil persistence is low; half-lives range from malathion, 1–2 wk, to parathion, diazinon, and azinphos-methyl, 3–6 mo. However, the organophosphorus insecticides are general biocides that are toxic to nearly all animal organisms. Oral LD₅₀ to mallard and pheasants are phorate 0.62, 7.1; parathion 2.0, 12; monocrotophos 4.8, 2.8; diazinon 3.5, 4.3; fenthion 5.9, 18; chlorpyrifos 75, 12; and azinphos-methyl 136, 75, respectively. These compounds are highly toxic to fish and LC₅₀ values for bluegill and trout (ppm) are parathion, 0.10, 0.047; fenthion, 1.4, 0.93; diazinon, 0.052, 0.38; azinphos-methyl: 0.005, 0.004; and malathion: 0.11, 0.13, respectively, and phorate (bluegill) 0.0055. The organophosphorus insecticides are highly toxic to bees and to beneficial parasites and predators. Great care should be taken in handling, applying, and storing these insecticides as the majority of insecticide poisonings throughout the world result from the use of the highly toxic OP compounds. It is essential to wear protective clothing including mask, goggles, and rubber gloves when handling these materials. In cases of severe exposure and incipient cholinergic symptoms, atropine should be administered as a specific antidote. Additional therapy with 2-pyridinealdoxime methiodide (2-PAM or pralidoxime) is of value in reversing the phosphorylation of AChE. The user of OP insecticides can enhance safety by choosing P=S esters in preference to P=O esters and *O,O*-dimethyl esters in lieu of *O,O*-diethyl esters.

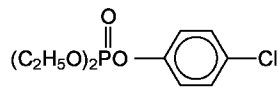
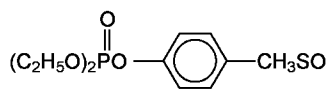
Carbamate Insecticides. These are structurally optimized derivatives of the unique plant alkaloid physostigmine [57-47-6], a cholinergic drug isolated in 1864 from *Physostigma venenosum* (see ALKALOIDS) (17,24,35–39). The carbamates may be considered synthetic derivatives of the synaptic neurotransmitter acetylcholine, with very low turnover numbers. The *N,N*-dimethylcarbamates of heterocyclic enols (36) and the *N*-methylcarbamates of a variety of substituted phenols (35) with a wide range of insecticidal activity were described in 1954 (35). The latter are the most widely used carbamate insecticides, and the *N*-methylcarbamates of oximes have subsequently been found to be effective systemic insecticides.

The carbamate insecticides represent the third principal group of synthetic organic insecticides developed after World War II. They are highly biodegradable and, although compounds such as carbofuran and aldicarb are highly toxic to mammals, their toxic action as inhibitors of acetylcholinesterase is rapidly reversible. Their use as insecticides has ameliorated both the environmental contamination experienced with the organochlorines and the hazards associated with the organophosphates, therefore carbamates have been widely used for foliar applications to field, vegetable, and fruit crops, as soil insecticides, to protect stored products, for veterinary hygiene, and with propoxur and bendiocarb as residual insecticides for malaria control in human habitations. Approximately 11,000 t of carbamates (active ingredients) were applied to U.S. farms in 1976 and worldwide application is estimated to exceed 35,000 t.

Carbaryl [63-25-2], 1-naphthyl *N*-methylcarbamate (**111**) (mp 142°C, *d* 1.232 g cm⁻³, vp 0.67 Pa at 20°C), is ca 95% pure as the technical grade and is soluble in water to 120 mg L. Carbaryl has a rat oral LD₅₀ of 540 mg kg and a dermal LD₅₀ of >2000 mg kg. It is a broad-spectrum insecticide registered on more than 100 crops. Carbaryl is rapidly detoxified and eliminated in animal urine and is neither concentrated in fat nor secreted in the butterfat of milk, thus it is favored for application to food crops.

Carbofuran [1563-66-2], 2,3-dihydro-2,2-dimethyl-7-benzofuranyl *N*-methylcarbamate (**112**) (mp 150–152°C), is soluble in water to 0.7 g L. The rat oral and dermal LD₅₀s are 4.8 and >10,000 mg kg, respectively. Carbofuran is a broad-spectrum soil insecticide.

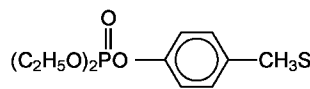
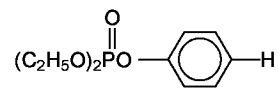




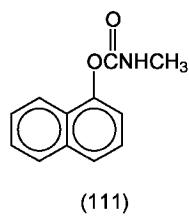
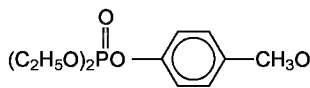
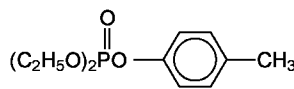
Propoxur [114-26-1], 2-isopropoxyphenyl *N*-methylcarbamate (**113**) (mp 91°C, vp 0.84 mPa), is soluble in nonpolar solvents and in water to 2 g L. The rat oral and dermal LD₅₀s are 83 and >1000 mg kg, respectively. Propoxur is used for the control of household insects, especially cockroaches, bedbugs, and mosquitoes, and is a replacement for DDT in malaria control.

Dioxacarb [6988-21-2], 2-(1,3-dioxolan-2-yl)-phenyl *N*-methylcarbamate (**114**) (mp 114–115°C, vp 50 μPa at 20°C), is water soluble to 6 g L. The rat oral and dermal LD₅₀ values are 125 and 3000 mg kg, respectively. Dioxacarb is used to control cockroaches and household and stored product pests.

Bendiocarb [22781-23-3], 2,2-dimethyl-1,3-benzodioxol-4-ol *N*-methylcarbamate (**115**) (mp 129–130°C, vp 0.7 mPa at 25°C), is water soluble to 0.04 g L. The rat oral and dermal LD₅₀s are 179 and 1000 mg kg, respectively. Bendiocarb is used to control cockroaches, household pests, and soil insects.

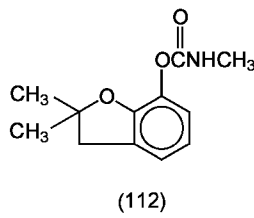


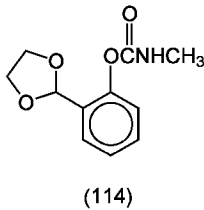
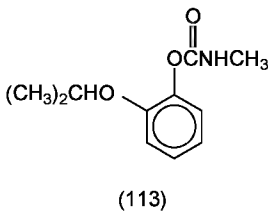
Mexacarbate [315-18-4], 4-dimethylamino-3,5-dimethylphenyl *N*-methylcarbamate (**116**) (mp 85°C), is soluble in water to 0.1 g L. The rat LD₅₀s are 37, 25 (oral) and 1500 (dermal) mg kg. It has been used to control caterpillars attacking forest trees. Aminocarb [2032-59-9], 4-dimethylamino-3-methylphenyl *N*-methylcarbamate (mp 93°C), is very closely related to mexacarbate and has rat LD₅₀s of 40, 38 (oral) and 280, 320 (dermal) mg kg. It is particularly effective against lepidoptera as well as snails and slugs. Methiocarb [2032-65-7], 4-methylthio-3,5-dimethylphenyl *N*-methylcarbamate (**117**) (mp 121°C, vp 15 mPa at 25°C), is soluble in water to 27 mg L. The rat LD₅₀s are 70, 60 (oral) and >2000 (dermal) mg kg. It is used to control insect pests of fruits and vegetables and snails and slugs. Formetanate [22259-30-9], 3-dimethylaminomethyleneiminophenyl *N*-methylcarbamate (**118**), is formulated as the water-soluble hydrochloride (vp 16 μPa at 25°C). The rat LD₅₀s are 15, 26 (oral) and >5000 (dermal) mg kg.



Isoprocarb [2631-40-5] is 2-isopropylphenyl *N*-methylcarbamate (**119**) (mp 93°C, vp 0.38 mPa at 20°C). The rat oral LD₅₀s are 403, 485 mg kg. It is used to control insect pests of rice and cacao. Promecarb [2631-37-0] is 3-methyl-5-isopropylphenyl *N*-methylcarbamate (mp 93°C, vp 4 mPa at 25°C) and is soluble in water to 91 mg L. The rat oral LD₅₀ is 74 mg kg. It is used to control orchard pests. Trimethacarb [2655-15-4] is 3,4,5-trimethylphenyl *N*-methylcarbamate (mp 117–119°C, vp 6.8 mPa at 25°C). The rat oral LD₅₀ is 178 mg kg. It is used as a soil insecticide.

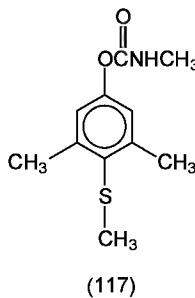
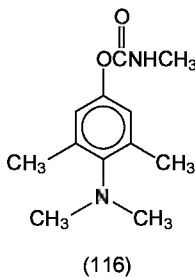
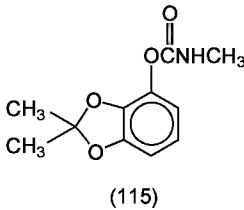
Ethiofencarb [29973-13-5] is 2-[2'-(methylthio)methyl]phenyl *N*-methylcarbamate (**120**) (*d* 1.147, vp 13 mPa at 30°C), soluble in water to 1.8 g L. The rat LD₅₀s are 411, 499 (oral) and 1150 (dermal) mg kg. It is a systemic insecticide effective against aphids. Etrofol [3942-54-9] is 2-chlorophenyl *N*-methylcarbamate (**121**) (mp 87°C). The rat oral LD₅₀ is 648 mg kg. It is used to control leafhoppers and plant hoppers in rice.





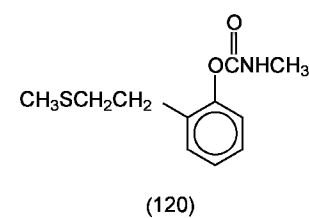
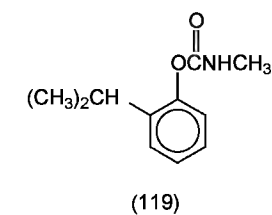
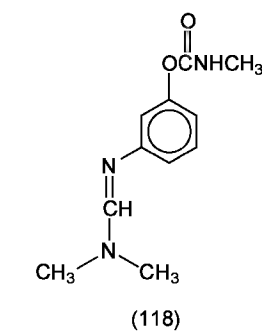
Other substituted phenyl *N*-methylcarbamates that have been used for insect control include metalkamate [8065-36-9], a 1:4 mixture of *m*-(1-ethylpropyl)-phenyl *N*-methylcarbamate [672-04-8] and *m*-(1-methylbutyl)-phenyl *N*-methylcarbamate [2282-34-0], rat oral LD₅₀s 87, 170 mg kg, used as a soil insecticide; MTMC [1129-41-5], 3-methylphenyl *N*-methylcarbamate (mp 76°C), rat oral LD₅₀ 268 mg kg, used for rice pests; MPMC [2425-10-7], 3,5-dimethylphenyl *N*-methylcarbamate (mp 79°C), rat oral LD₅₀ 380 mg kg, used for rice pests; TBPMC [780-11-0], 3-*tert*-butylphenyl *N*-methylcarbamate (mp 140°C), mouse oral LD₅₀ 470 mg kg, for rice pests; and butacarb [2655-19-8], 3,5-di-*tert*-butylphenyl *N*-methylcarbamate (mp 98°C), rat oral LD₅₀ >4000 mg kg, for sheep blowfly larvae control.

Dimetilan [644-64-4], 2-(*N,N*-dimethylcarbamoyl)-3-methylpyrazol-5-yl *N,N*-dimethylcarbamate (**122**) (mp 68°C) has rat LD₅₀s of 25, 64 (oral) and 600 (dermal) mg kg. It is used in sugar bait for fly control. Pyrimicarb [23103-98-2], 2-(dimethylamino)-5,6-dimethyl-4-pyrimidinyl *N,N*-dimethylcarbamate (**123**) (mp 90°C, vp 4 mPa at 30°C), is soluble in water to 2.6 g L. The rat LD₅₀s are 147 (oral) and 500 (dermal) mg kg. It is a systemic aphicide used primarily on grain crops.

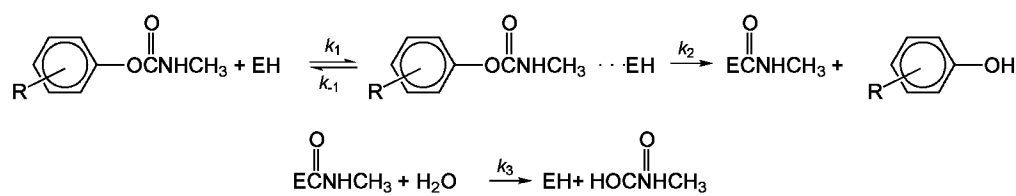


Aldicarb [116-06-3], *O*-(methylcarbamoyl)-2-methyl-2-methylthiopropionaldehydeoxime (**124**) (mp 99–100°C, vp 13 mPa at 20°C), is soluble in water to 6 g L. The rat oral LD₅₀s are 0.8, 0.65 (oral) and 2, 2.5 (dermal) mg kg. Aldicarb is a broad-spectrum systemic insecticide used for seed and soil treatment and as a nematocide. Aldicarb readily oxidizes to the corresponding sulfone, aldoxycarb [1646-88-4] (mp 99°C, vp 12 mPa at 25°C), which is soluble in water to 10 g L. The rat oral LD₅₀ is 27 mg kg. Aldoxycarb is also a systemic insecticide.

Methomyl [16752-77-5], *S*-methyl, *N*-[(methylcarbamoyl)oxy]thioacetimidate (**125**) (mp 78–79°C, vp 7 mPa at 20°C), is soluble in water to 58 g L. The rat LD₅₀s are 17, 26 (oral) and 15,000 (dermal) mg kg. Methomyl is a broad-spectrum insecticide of relatively short persistence. Oxamyl [23135-22-0], *S*-methyl-1-(dimethylcarbamoyl)-*N*-[(methylcarbamoyl)oxy]thioformimidate (**126**) (mp 108–110°C, vp 20 mPa at 25°C), is soluble in water to 280 g L. Oxamyl is a systemic insecticide and nematocide. Thiofanox [39196-18-4], 3,3-dimethyl-1-(methylthio)-2-butanone *O*-[(methylamino)carbonyl]oxime (**127**) (mp 57°C, vp 22.6 mPa at 25°C), is soluble in water to 5.2 g L. The rat oral LD₅₀ is 8.5 mg kg. Thiofanox is a systemic insecticide effective against red spider mites.



Mode of Action. All of the insecticidal carbamates are cholinergic, and poisoned insects and mammals exhibit violent convulsions and other neuromuscular disturbances. The insecticides are strong carbamylating inhibitors of acetylcholinesterase and may also have a direct action on the acetylcholine receptors because of their pronounced structural resemblance to acetylcholine. The overall mechanism for carbamate interaction with acetylcholinesterase is analogous to the normal three-step hydrolysis of acetylcholine; however, k_3 is much slower than with the acetylated enzyme.



The importance of structural complementarity is demonstrated by the fivefold greater toxicity of 1-2-*sec*-butylphenyl *N*-methylcarbamate [75863-94-4] over the corresponding *D*-isomer, and by the greatly reduced toxicity of 4-isopropylphenyl *N*-methylcarbamate [64-00-6] as compared to the 2-isopropyl [114-26-1] and 3-isopropyl isomers [4089-99-0]. Detoxication of carbamate insecticides occurs *in vivo* through microsomal oxidase action, producing *N*-demethylation of the carbamate nitrogen, side-chain oxidation, and ring hydroxylation. Methylenedioxyphenyl synergists are therefore highly effective with carbamates.

Environmental. The *N*-methylcarbamates generally are biodegradable and of low soil persistence with half-lives for carbaryl and aldicarb of 1–2 weeks and of carbofuran of 1–4 months. Certain carbamates are highly toxic to birds with oral LD₅₀s for mallard, eg, pheasant, in mg/kg: carbofuran, 0.40, 4.2; mexacarbate, 3.0, 4.5; and methomyl, 16, 15; compared to carbaryl >2000. Fish toxicity of carbamates is generally low, but these compounds are extremely toxic to bees. In cases of human poisoning, atropine is a specific antidote.

Insect Growth Regulators. These compounds (40–45), unlike most conventional insecticides, interfere with biochemical processes that are unique to arthropods; eg, molting, ecdysis, and formation of the chitinous exoskeleton. Therefore, they are selective insecticides with very low mammalian toxicity.

Juvenoids. The discovery of the insect juvenile hormone (JH) or neotenin of the corpus cardiacum, methyl (*E,E*)-*cis*-10,11-epoxy-7-ethyl-3,11-dimethyl-tridecadi-2,6-enoate (128), initiated the study of thousands of analogues produced by structural optimization and several have been produced as practical insecticides (24,40–43). These act at the receptor sites for neotenin and thus affect embryogenesis, metamorphosis, reproduction, diapause, and other critical developmental processes. Neotenin is insecticidal when applied exogenously to insect eggs and last instar larvae and interferes with the development of physiologically competent adults. The structurally optimized analogues are more lipophilic and persistent. As shown in Table 5, they are often highly specific in insecticidal action. Their value in insect control lies in their extremely low toxicity to mammals and drawbacks for commercial pest control are very slow toxic action and general lack of persistence.

Table 5. Insecticidal Effectiveness of Some Juvenoids

Compound ^a	CAS Registry Number	LD ₅₀ values				
		<i>Tenebrio molitor</i> , μg pupa	<i>Galleria melonella</i> , μg pupa	<i>Aedes aegypti</i> , ppm	<i>Musca domestica</i> , μg pupa	<i>Heliothis virescens</i> , ppm
neotenin	[32766-80-6]	0.70	0.060	0.15	>100	24
epiphenonane	[57342-02-6]	0.0024	0.037	0.057	54	2.2
hydroprene	[41096-46-2]	0.25	0.040	0.0078	18	0.30
methoprene	[40596-69-8]	0.0040	5.7	0.00017	0.0035	0.77
fenoxycarb	[79127-80-3]		0.00034	0.000022	0.040	0.018

^a See Figure 1.

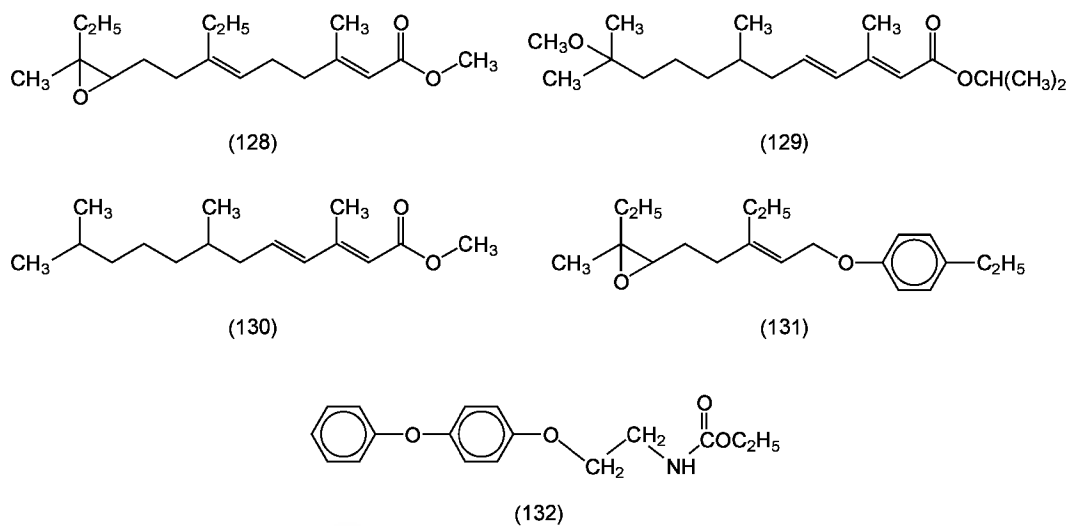


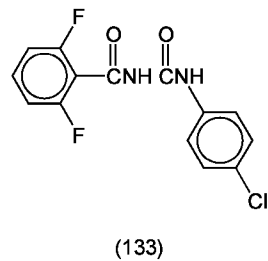
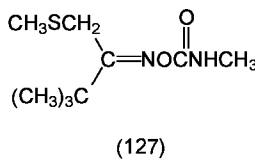
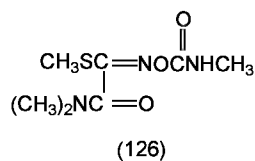
Fig. 1. First- and second-generation juvenoids. See Table 5.

Methoprene and hydroprene are first-generation juvenoids that incorporate minor structural optimization of neotenin to increase persistence. Methoprene, 1-isopropyl (*E,E*)-11-methoxy-3,7,11-trimethyl dodecadi-2,4-enoate (**129**) (bp 100° C 6.7 Pa, vp 3.5 mPa at 25°C), is soluble in water to 1.4 mg L. The rat oral LD₅₀ is >34,000 mg kg. Methoprene has been used as a mosquito larvicide, in baits for ant control, and as a cattle feed-through treatment for hornfly control. Hydroprene, methyl (*E,E*)-3,7,11-trimethyl-dodecadi-2,4-enoate (**130**) (bp 174°C at 2.5 kPa, vp 2.5 mPa at 25°C), is soluble in water to 0.54 mg L. The rat oral LD₅₀ is >34,000 mg kg. Hydroprene is especially effective against aphids and cockroaches.

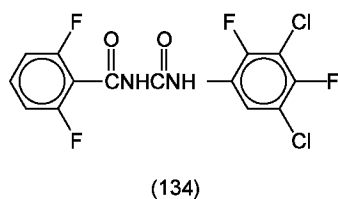
Second-generation juvenoids incorporate more substantial structural departures from neotenin and are more resistant to metabolic and environmental degradation. Epiphenonane, 2-ethyl-3-[3-ethyl-5-(4-ethylphenoxy)-pent-3-en-yl] 2-methyloxirane (**131**), has a rat oral LD₅₀ of 4000 mg kg. It and similar juvenoids are used in China and Japan to prolong the last larval instar of the silkworm so that silk production is increased 10–15%. Fenoxycarb, ethyl [2-(4-phenoxyphenoxy)ethyl] carbamate (**132**) (mp 53°C, vp 0.0078 mPa at 20°C), is soluble in water to 6 mg L. The rat oral LD₅₀ is >16,800 mg kg. Fenoxycarb has a wide spectrum of activity, interfering with the developmental processes of fleas, cockroaches, and ants.

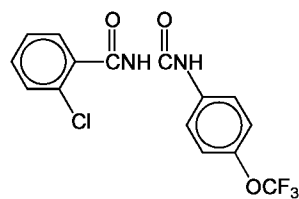
Chitin Synthesis Inhibitors. These are insect growth regulators that prevent the formation of the insect chitinous exoskeleton and thus produce a critical biochemical lesion during hatching, ecdysis, or pupation (44). These complex biochemical and physiological processes are unique to arthropods; therefore, the benzoyl phenyl ureas are highly specific insecticides (44).

Diflubenzuron [35367-38-9], 1-(4-chlorophenyl)-3-(2,6-difluorobenzoyl) urea (**133**) (mp 239°C, vp 0.033 mPa at 50°C), is soluble in water to 300 µg L. The rat LD₅₀ value is >4640 mg kg. Teflubenzuron [83121-18-10], 1-(2,4-difluoro-3,5-dichlorophenyl)-3-(2,6-difluorobenzoyl) urea (**134**) (mp 222°C, vp 0.8 mPa at 20°C), is soluble in water to 19 µg L. The rat LD₅₀ values are >5000 (oral) and >2000 (dermal) mg kg.



Triflumaron [64628-44-0], 1-(4-trifluoromethoxyphenyl)-3-(2-chlorobenzoyl) urea (**135**) (mp 195°C, vp 40 mPa at 20°C), is soluble in water to 25 µg L. The rat LD₅₀ values are >5000 (oral and dermal) mg kg. Hexaflumaron [86479-06-3], 1-[3,5-dichloro-4-(1,1,2,2-tetrafluoroethoxy)-phenyl]-3-(2,6-difluorobenzoyl) urea (**136**) (mp 202°C, vp 0.059 mPa at 25°C), is soluble in water to 27 µg L. The rat LD₅₀ values are >5000 (oral, dermal) mg kg.

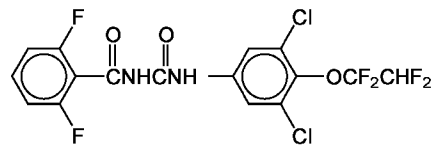




(135)

Flucycloxuron [94050-53-0], 1-[α-(4-chloro-α-cyclopropylbenzylideneaminoxy)-4-tolyl]-3-(2,6-difluorobenzoylurea (**137**) (mp 144°C, vp 4.4 mPa at 20°C), is soluble in water to 1 μg L. The rat LD₅₀ values are >5000 (oral) and >2000 (dermal) mg per kg. These insecticides are slow-acting ovicides and larvicides with little contact action. They are especially effective against Lepidoptera, Coleoptera, and Diptera. Their mode of action is believed to be as inhibitors of the enzyme chitin synthetase, which polymerizes uridinediphosphoacetylglucoseamine to chitin.

Azadirachtin [11141-17-6] (**138**) and related limonoid triterpenoids present in the seeds of the neem tree, >Azadirachta indica (Meliaceae), to about 0.2–0.4% are insect growth regulators that inhibit molting and produce sterility in a wide range of insect species (45). Neem extracts are registered for use as insecticides on a variety of food crops, and for control of insects infesting greenhouse plants and turf.



(136)

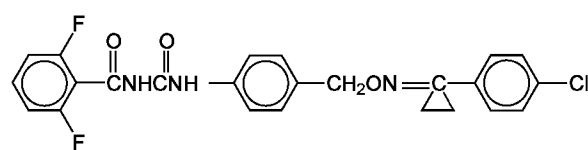
Acaricides

Chemicals that are especially effective in controlling the mites and ticks of the order Acarina are acaricides (1). Most of the insecticidal chemicals discussed, with the exception of the organophosphorus insecticides, are not of practical value as acaricides. Applications of many of the chlorinated hydrocarbon insecticides have no effect on the phytophagous spider mites but often kill their predators resulting in abnormally large populations of the spider mites. A number of acaricidal chemicals have come into widespread use that have almost specific toxicity to the mites but are inactive against insects. In general, these acaricides are highly stable compounds with comparatively prolonged residual action and low mammalian toxicity. Certain of the compounds described are effective only as ovicides, killing the eggs and sometimes the newly emerged nymphs, and others are active against all stages of mites. These acaricides exhibit a considerable degree of specificity for various species of acarina and are most useful for the phytophagous Tetranychidae and Eriophyidae. Other acaricides that repel or kill mites and ticks that attack humans and animals are described (see REPELLENTS).

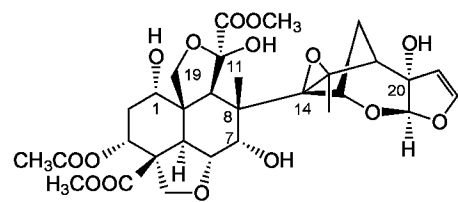
Chlorfenethol [80-08-6], 1,1-bis(*p*-chlorophenyl)ethanol, R₂H (**139**), is a white solid (mp 70°C). The compound is readily dehydrated upon heating or in the presence of strong acids and forms the inactive 1,1-bis(*p*-chlorophenyl)ethylene. Chlorfenethol is active against all stages of mites. It has an oral LD₅₀ to the rat of ca 200 mg kg.

Chlorobenzilate [510-15-6], ethyl *p,p'*-dichlorobenzilate (**140**), is a yellowish viscous oil (bp 141–142°C at 8 Pa). The technical material (*d* 1.281) contains ca 90% of the active compound, is insoluble in water, and is soluble to more than 40% in deodorized kerosene, benzene, and methyl alcohol. Chlorobenzilate is hydrolyzed in alkali and in strong acids to the inactive *p,p'*-dichlorobenzilic acid and ethanol. The compound is active against all stages of mites and has an oral LD₅₀ to the rat of 700 mg kg.

Dicofol [54532-36-4], 1,1-bis(*p*-chlorophenyl)-2,2,2-trichloroethanol, R₂Cl (**139**), is a white crystalline solid (mp 79°C). This compound is insoluble in water and soluble in organic solvents, and in the presence of alkali forms the inactive *p,p'*-dichlorobenzophenone and chloroform. Dicofol is a long-lasting acaricide and is active against all stages of mites. The rat oral LD₅₀s are 809, 684 mg kg.



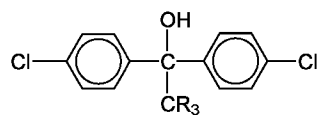
(137)



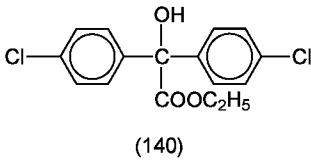
(138)

Tetradifon [116-29-0], 2,4,5,4'-tetrachlorodiphenyl sulfone (**141**), is a crystalline solid (mp 148°C). It is soluble in water to 200 mg L. Tetradifon is stable to the action of acid and alkalis, light and temperature, and has a very prolonged residual action. It is active against all stages of mites and has an oral LD₅₀ to the rat of >14,700 mg kg.

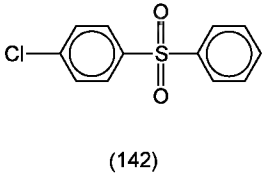
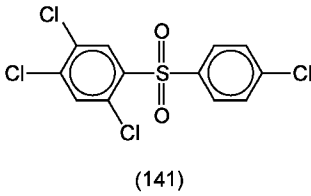
Sulphenone [80-30-2], *p*-chlorophenyl phenyl sulfone (**142**), is a white solid (mp 98°C). The technical material consists of ca 80% of this compound, with small amounts of *o*- and *m*-isomers, bis(*p*-chlorophenyl) sulfone, and diphenyl sulfone. Sulphenone is effective against all stages of mites. Its oral LD₅₀ to the rat is >2000 mg.



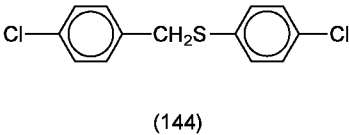
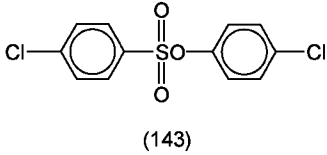
(139)



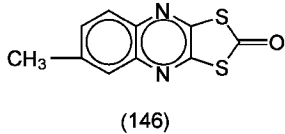
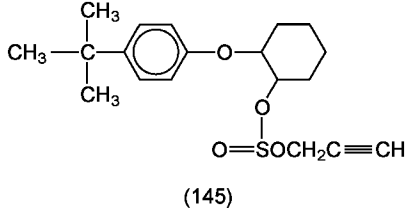
Ovex [80-33-1], *p*-chlorophenyl *p*-chlorobenzenesulfonate (**143**), is a white solid (mp 86.5°C). Ovex is effective only as an ovicide. It has an oral LD₅₀ to the rat of 2000 mg kg. Two analogues of ovex have similar ovicidal properties and, like ovex, are hydrolyzed in alkali to form the phenol and benzenesulfonate salt: Genite [97-16-5], 2,4-dichlorophenyl benzenesulfonate (mp 45–47°C, vp 36 mPa at 30°C), is the technical product and is about 97% pure having an oral LD₅₀ to the rat of 1400 mg kg; fenon [80-38-6] is *p*-chlorophenyl benzenesulfonate (mp 61–62°C, *d* 1.33 rat oral LD₅₀ 1600 mg kg). Chlorbenside [103-17-3] is *p*-chlorobenzyl *p*-chlorophenyl sulfide (**144**) (mp 74°C, vp 35 mPa at 20°C). The technical product contains ca 90% *p,p'*-isomer, 5% *o,p'*-isomer, and 2.5% *m,p'*-isomer. Chlorbenside is unaffected by reduction and by acid and alkaline hydrolysis, but it is readily oxidized to *p*-chlorobenzyl *p*-chlorophenyl sulfoxide [7047-28-1] (mp 125°C) and more slowly to *p*-chlorobenzyl *p*-chlorophenyl sulfone [74512-22-4] (mp 150°C). These reactions occur on the leaf surface, and the oxidation products are acaricidal but do not penetrate locally into the leaf tissue as does chlorbenside. Chlorbenside is active only against eggs and immature mites. The oral LD₅₀ to the rat is >10,000 mg kg.

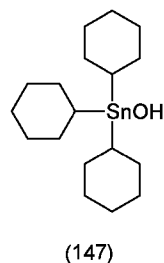


Propargite [2312-35-8], 2-(*p*-*tert*-butylphenoxy)cyclohexyl 2-propynyl sulfite (**145**), is a dark oil (*d* 1.1), is insoluble in water, and hydrolyzes in alkaline solutions. The rat oral LD₅₀ is 2200 mg kg. Propargite is a widely used acaricide on fruits, vegetables, and row crops. Oxythioquinox [2439-01-2], 6-methyl-2,3-quinoxalinedithiol cyclic carbonate (**146**) (mp 172°C, vp 26 μPa at 20°C), is insoluble in water and soluble in organic solvents with rat oral LD₅₀s of 1800, 1100 mg kg. It is used as an acaricide for tree fruits.



Cyhexatin [13121-70-5], tricyclohexylhydroxystannane (**147**) (mp 195°C), rat oral LD₅₀ 540 mg kg, and fenbutatin oxide [13356-08-6], hexakis-(2-methyl-2-phenylpropyl)distannoxane (**148**) (mp 138°C), rat oral LD₅₀ 2630 mg kg, are two novel tin acaricides used on deciduous fruits. They are inhibitors of oxidative phosphorylation.

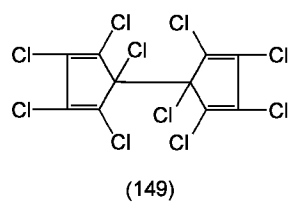
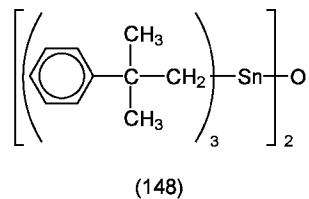




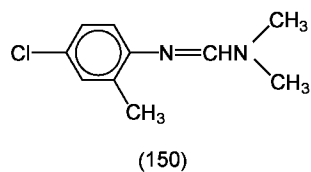
Dienochlor [2227-17-0], bis-(pentachloro-2,4-cyclopentadien-1-yl) (149) (mp 122°C), is used as an acaricide for greenhouse and ornamental crops. The rat oral LD₅₀ is 3160 mg kg.

Miscellaneous Insecticides

Formamidines. These are competitive agonists of octopamine or 1-(*p*-hydroxyphenyl)-2-aminoethanol [104-14-3], an insect neurotransmitter (24). Chlordimeform [6164-98-3] is *N'*-(4-chloro-2-methylphenyl)-*N,N*-dimethylformamidine (150) (mp 35°C, *d* 1.105, vp 0.045 mPa at 20°C). The rat oral LD₅₀ range is 170–340 mg kg. It is used as an acaricide and as an ovicide to control bollworms and the rice stem borer. However, it is a bladder carcinogen, and its use is restricted. Amitraz [33089-61-1] is *N'*-2,4-dimethylphenyl-*N*-(*N'*-2,4-dimethylphenyl)-imino *N*-methylmethanimidamide (151) (mp 86°C), and is soluble in water to 1 mg L. The rat oral LD₅₀ is 600–800 and the dermal LD₅₀ is 1600 mg kg. It is used to control spider mites, pear psylla, and ticks.



Avermectins and Ivermectin. The avermectins are pentacyclic lactones isolated from fermentation products of *Streptomyces avermitilis*, and ivermectin is a semisynthetic chemical, 22,23-dihydroavermectin B₁ (46). Ivermectin is effective in very low doses for the control of red spider mites on deciduous fruits, in baits for the control of imported fire ants, and as a parasiticide for *Onchocerca volvulus* in humans and for cattle grubs. These insecticides appear to function as agonists for the neuroinhibitory transmitter γ-aminobutyric acid (GABA) (see ANTIPARASITIC AGENTS, AVERMECTINS). Hydramethylnon [67485-29-4] is tetrahydro-5,5-dimethyl-2-(1*H*)-pyrimidinone [bis-1,5-(4-trifluoromethylphenyl)-3-penta-1,4-dienylidene] hydrazone (152) (mp 189°C). It is a slow-acting stomach poison used in baits and traps to control ants and cockroaches. Its mode of action is inhibition of mitochondrial electron transport.



Sulfluramid or *N*-ethyl perfluorooctanesulfonamide [41 51-50-2], C₈F₁₇SO₂NHC₂H₅, is a slow-acting stomach poison used in baits for the control of ants and cockroaches. It acts as an uncoupler of oxidative phosphorylation.

Petroleum Oils. When satisfactorily stable kerosene–soap–water emulsions were produced in 1874, dormant (winter) oil sprays became widely used to control scale insects and mites (1). The first commercial emulsion or miscible oil was marketed in 1904 and by 1930 highly refined neutral or white oils, free from unsaturated hydrocarbons, acids, and highly volatile elements, were found to be safe when applied to plant foliage, thus greatly enlarging the area of usefulness of oil sprays (see PETROLEUM).

Petroleum oil sprays are used as insecticides for dormant sprays in the control of scale insects, mites, and insect eggs; summer foliage sprays for aphids, mealybugs, mites, thrips, psyllids, whiteflies, and scale insects; livestock sprays for the control of lice, fleas, and mites; and mosquito larvicides. They also are used as carriers for contact insecticides to increase their effectiveness. In order to understand the effects of spray oils on insects and plants, volatility or distillation range must be considered. The lower the volatility and the higher the boiling point, other properties being equal, the more effective the oil is in killing insects. For dormant spraying, the oil should distill ca 90% of its volume from 310–371°C and not over 2% at 110°C for 4 h. Unfortunately, the heavier oils are more toxic to plants so it is important to find the lightest oil that will kill the insect pest or the heaviest that can be safely used on the plant to be sprayed.

The viscosity of the spray oil, as measured by the Saybolt test, also determines its safety on plants. Other properties being equal, oils of low viscosity are safer to use on foliage than those of high viscosity. For dormant sprays on deciduous trees, oils with viscosities between 100 and 200 Saybolt universal seconds (SUs) at 37.8°C are considered satisfactory. A lower range is often used in colder and a higher range in warmer areas.

The degree of refinement of the spray oil is important because the presence of impurities, eg, unsaturated and aromatic hydrocarbons, which are chemically reactive and readily oxidized, causes the oil to become turbid and acidic in reaction. For dormant spraying, 65–75% of these impurities must be removed and for foliage spraying, 85–100% removal is necessary. The refining is accomplished chiefly by treatment with sulfuric acid and by washing to remove the resulting sludge. For any oil of unknown purity, the degree of refinement may be determined by treating a sample with sulfuric acid. If the unsaturated hydrocarbons have already been removed, the acid treatment does not remove any portion of the oil and it is said to be 100% unsulfonatable. Typical specifications for spray oils are shown in Table 6.

Table 6. Specifications for Spray Oils

Grade	Distillation at 336°C, ° o	Unsulfonated residue (UR), ° o	Viscosity at 37.8°C, SUs
dormant oil		90–92	90–120
foliage oil			
light	64–79	90	55–65
light–medium	52–61	92	60–75
medium	40–49	92	70–85
heavy–medium	28–37	92	80–95
heavy	10–25	94	90–105

Fumigants. Fumigants are chemicals that are distributed through space as gases and therefore, at a given temperature and pressure, must exist in the gaseous state in sufficient concentration to be lethal to the insect pest (1,47). This physical requirement greatly limits the number of insecticides that may be usefully employed as fumigants. Compounds boiling at about room temperature, eg, hydrogen cyanide, methyl bromide, and ethylene oxide, are the most useful general fumigants. For soil fumigation, however, the slower release of vapors from substances, eg, ethylene dibromide and β,β'-dichlorodiethyl ether that boil as high as 180°C, has proven effective. Other organic toxicants of relatively high vapor pressure, eg, naphthalene and *p*-dichlorobenzene, sublime readily enough to have special uses as fumigants; and contact insecticides, eg, lindane, dichlorvos, and mevinphos, may kill insects by vapor action under certain circumstances. The fumigant action of the less easily vaporizable insecticides is enhanced by the use of atomization, volatilization by heat, or burning in pyrotechnic mixtures. The true insecticidal fumigants are described in Table 7. The principal areas for use are mills, warehouses, grain elevators, groceries, museums, etc; stored food products either in bulk or in small packages; habitations for control of structural and household pests and for human or animal ectoparasites; soil fumigation for grubs, wireworms, ants, nematodes, etc; and living plants in greenhouses, nursery stock, or citrus and other tree crops.

Table 7. Properties of Fumigants

Name	CAS Registry Number	Formula	Boilin g point, °C	Specific gravity		Vapor pressu re ^{a,b} kPa	Solubility, ^a g 100 mL H ₂ O	Flamma bility in air, ^c vol % _o	TLV safe limit, ^d ppm	Uses
				Liquid <i>d</i> ₄ ^o , ^a g mL	Gas (air = 1) ()					
acrylonitrile	[107-13-1]	CH ₂ =CHCN	78	0.797	1.8	11		3	20	mills, commodities
carbon disulfide	[75-15-0]	CS ₂	46.3	1.263	2.6	41.9	0.22	1	10	household
chloropicrin	[76-06-2]	Cl ₃ CNO ₂	112	1.651	5.7	2.7	0.19	nf	0.1	soil, grain
1,1-dichloro-1-nitroet hane	[594-72-9]	CH ₃ CCl ₂ NO ₂	124	1.415		2.25	0.25		10	grain, stored products
1,2-dichloropropane	[78-87-5]	CH ₂ ClCHClCH ₃	95.4	1.159		28	0.27		50	soil
<i>trans</i> -1,3-dichloro-pro pene	[542-75-6]	ClCH=CHCH ₂ Cl	111	1.224		2.47	0.28	f	1	soil
ethylene chlorobromide	[107-04-0]	ClCH ₂ CH ₂ Br	107	1.689		5.3	0.7 ³⁰	nf		soil, commodities
ethylene dichloride	[107-06-2]	ClCH ₂ CH ₂ Cl	83.5	1.257	3.4	10	0.87	6		grain, soil household
ethyl formate	[109-94-4]	HCOOC ₂ H ₅	54	0.917			10		100	dried fruits
ethylene oxide	[75-21-8]	(CH ₂) ₂ O	10.7	0.887 ⁷	1.5	146	∞	3	1	packaged foods
hydrogen cyanide	[74-90-8]	HCN	26	0.688	0.9	84.0	∞	6	10	general
methyl bromide	[74-83-9]	CH ₃ Br	4.5	1.732 ⁰	3.3	189.3	1.34 ²⁵	13.5	5	general
methyl formate	[107-31-3]	HCOOCH ₃	32	0.974		83.2	30		100	dried fruits
naphthalene	[91-20-3]	C ₁₀ H ₈	218		4.4	0.01	0.003	f	10	fabric pests, greenhouse
			(mp 80)							
<i>p</i> -dichlorobenzene	[106-46-7]	C H ₄ Cl ₂	173.4		5.1	0.1	0.008 ²⁵	nf	75	fabric pests
			(mp 53)							
phosphine	[7803-51- 2]	PH ₃	87.4	0.746 ⁹⁰	1.2			2	0.3	grain
sulfuryl fluoride	[2699-79- 8]	SO ₂ F ₂	55.2	1.342 ²⁵	3.5	1601.3	0.075 ²⁵	nf	5	structural pests
trichloroacetoneitrile	[545-06-2]	Cl ₃ CCN	85	1.44 ²⁵						grain
trichloroethylene	[79-01-6]	ClCH=CCl ₂	86.7	1.470 ¹⁵	4.5	9.7	insol	nf	50	grain

^b To convert kPa to mm Hg, multiply by 7.5.

^a At 20°C unless otherwise indicated by superscript, °C.

^c nf nonflammable; f flammable.

^d Threshold limit value, time-weighted average.

The various fumigants often exhibit considerable specificity toward insect pests, as shown in Table 8. The proper choice for any control operation is determined not only by the effectiveness of the gas but by cost; safety to humans, animals, and plants; flammability; penetratability; effect on seed germination; and reactivity with furnishings. The fumigants may be used individually or in combination. Carbon tetrachloride has been incorporated with carbon disulfide, ethylene dichloride, or ethylene dibromide to decrease flammability, and carbon dioxide is used with ethylene oxide for the same purpose.

Table 8. Comparative Toxicity of Fumigants, LC₅₀ for 6 h, mg/L^a

--

Fumigant	Granary weevil, <i>Sitophilus granarius</i>	Drugstore beetle, <i>Stegobium paniceum</i>	Confused flour beetle, <i>Tribolium confusum</i>	Bean, weevil, <i>Acanthoscelides obtectus</i>	Saw-toothed grain weevil, <i>Oryzaephilus surinamensis</i>
acrylonitrile	2.0	1.7	3.0	1.1	0.8
carbon disulfide	43.0	42.0	75.0	29.0	40.0
chloropicrin	3.4	1.9	6.4	<1.5	<1.5
ethylene dibromide	3.0	2.8	3.4	10.2	0.9
ethylene dichloride	127.0	77.0	53.0	49.0	39.0
ethylene oxide	13.5	9.0	27.5	10.5	4.0
hydrogen cyanide	4.6	<0.4	0.8	0.9	<0.4
methyl bromide	4.8	4.4	9.2	4.2	4.4

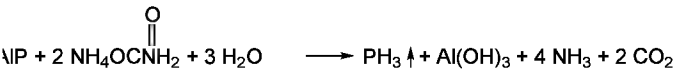
^a Ref. 47.

Dosage. The dosage of fumigant is commonly expressed as $\text{lb } 1000 \text{ ft}^3$ or in mg L . Successful fumigation results from the attainment of a critical (C_t) value (product of the concentration of gas in $\text{mg L} \times$ the exposure duration in h), ie, that which will kill at least 99% of the insect population. Within moderate limits, therefore, the longer the exposure, the lower the concentration of gas necessary. In practice, the exposure time is limited by the escape of gas from the enclosure and convenience in treatment; the minimum concentration is limited by the ability of the pest to detoxify the gas at low dosages as rapidly as it is sorbed. The dosage and the attainment of a critical C_t value are dependent on the rate of volatilization of generation of the fumigant and the nature of the commodity being fumigated. Some fumigants, eg, methyl bromide, are highly reactive with proteins and are rapidly sorbed by grain, flour, or seeds. From 10–32°C, the C_t value decreases by approximately one half for each 10°C increase in temperature.

Application. The fumigant is applied to an enclosure that is as gastight as possible. Low boiling fumigants, eg, hydrogen cyanide, methyl bromide, ethylene oxide, methyl formate, and sulfuryl fluoride, are obtained in cylinders of compressed or liquefied gas which is readily piped into the enclosure. Fumigants that are liquids at room temperature are volatilized by pouring onto cloths or spraying into the area to be treated. Elaborate metering mechanisms have been developed for spot applications of fumigants to moving streams of grain in elevators. Forced recirculation of the fumigant through the commodity being treated improves the distribution of the gas and often is used for the fumigation of stored grains.

Certain fumigants may be controlled more readily if generated from relatively inert precursors. Hydrogen cyanide is rapidly released by treatment of sodium, potassium, or calcium cyanides with acid, or slowly upon exposure to moisture. Small packages of sodium cyanide may be dropped into containers of dilute acid to provide precisely controlled dosages of hydrogen cyanide for household or warehouse fumigation. Calcium cyanide absorbed on inert earth or in pellets may be spread in a thin layer on greenhouse walks or on paper on household floors; the hydrogen cyanide gas is generated by the moisture in the air.

The extremely toxic and flammable gas phosphine is safely and conveniently generated for the fumigation of grain in sacks or bins from 3-g tablets containing aluminum phosphide and ammonium carbamate which produce 1 g of phosphine in the presence of moisture.



The organophosphorus ester dichlorvos is sufficiently volatile to be incorporated either in permeable plastic bottles or plastic strips which permit its controlled release for fumigation of cupboards and closets to control cockroaches, or as flea collars on pets.

Vacuum Fumigation. The molecular diffusion of a fumigant gas is inversely proportional to its molecular weight and is affected also by the number of collisions of the fumigant molecules with other gases present. The penetration and effectiveness of the fumigant may be improved greatly by using it in a partial vacuum. Vacuum fumigation of packaged foods, tobacco, spices, pharmaceuticals, and other commodities is carried out in steel chambers or vaults at a pressure of 2–23 kPa (15–175 mm Hg). The reduction in oxygen content makes the insects more susceptible to the effects of the fumigant and the addition of carbon dioxide frequently is used to stimulate the insects to respire at a higher rate and thus increases the sorption of the fumigant.

Tent Fumigation. Citrus and deciduous fruit trees have been fumigated for the control of scale insects for many years by hydrogen cyanide introduced under relatively gastight tents of 240–270 g m⁻² (7–8 oz yd²) army duck or nylon impregnated with vinyl chloride–vinyl acetate copolymer. Methyl bromide fumigation of buildings for the control of termites or powder post beetles, of mills or warehouses, or of bagged commodities for the control of stored-product insect pests, is effected by wrapping the structure in polyethylene or poly(vinyl chloride) plastic sheeting that is 0.102–0.152 mm thick, rolled and clamped at the edges, and sealed at the bottom by soil or sand bags.

Soil Fumigation. This procedure has become a standard agricultural practice for the destruction of soil-inhabiting insects, nematodes, and fungi prior to planting, in orchard and ornamental plantings, and in turf. The fumigant may be injected directly ahead of the plow or applied by more elaborate soil injectors. Small plots are treated by pouring the fumigant into holes punched in the soil at regular intervals and covering immediately. More volatile soil fumigants, eg, methyl bromide, are applied under plastic sheeting that is sealed at the edges with earth. The distribution of the soil fumigant involves vaporization outward from the point of application through air spaces in the soil and solution of the fumigant in soil water. The most efficient pattern for application of specific fumigants to various types of soils can be computed from the mathematical laws governing the diffusion of gases.

Environmental. The fumigants generally are highly reactive compounds that interact with vital biochemical processes within the target pest, usually by a bimolecular process, eg, alkylation. Their reactivity and high vapor pressure make it difficult to prevent widespread contamination of the surrounding air, water, and soil. Soil fumigation practiced with highly persistent, lipophilic compounds (see Table 7) at high dosages of 90–450 kg/ha (80–400 lb/acre) has contaminated the treated food commodity and water in shallow wells used for domestic supply. Fumigants, such as carbon tetrachloride, ethylene oxide, β,β' -dichlorodiethyl ether, ethylene dibromide, and 1,2-dibromo-3-chloropropane (DBCP) [96-12-8] are chemical carcinogens; the latter two compounds have produced azoospermia in workers exposed in both factory production and agricultural operations. For these reasons, all U.S. usage of DBCP was canceled in 1979.

Microbial Control. Insects are attacked by a multitude of pathogens (47–51). More than 450 viruses, 100 bacteria, 460 fungi, 250 protozoa, and 20 rickettsia are recognized as effective natural enemies. A number of these are adaptable for mechanical dissemination as microbial insecticides for the inoculation of insect populations, soils, fields, orchards, or forests with spores, virus suspensions, or microbial toxins. Such microbial insecticides are highly selective to insect pests, do not leave toxic residues on food crops, and are compatible with all other means of insect control. Thus they are increasingly favored for use in programs for integrated pest management (IPM). They have, however, the disadvantages of very short residual action due to inactivation by uv light and desiccation, extreme specificity that limits general applicability, slow lethal action, and requirement for precise timing of applications.

Insect pathogens are used in three different ways: (1) to maximize the spread of naturally occurring epizootics, (2) by introducing them into pest populations as permanent mortality factors, and (3) by application as microbial insecticides consisting of suspensions of spores, virus particles, or microbial toxins. Microbial insecticides have great potential usefulness in IPM programs. However, they presently comprise only about 1% of the world insecticide market, but their usage growth rate is presently about 10 times greater than that of conventional synthetic insecticides. Commercial microbial insecticides are listed in Table 9. Bacterial products are presently the most widely used because of ease of production. The varieties of *Bacillus thuringiensis* (*Bt*) which was first commercialized in 1957, display an increasing span of pest control activity which is largely due to the formation of crystalline delta-endotoxin. The available products are highly standardized and extensive efforts are being made to improve their properties by microencapsulation, and by a variety of molecular techniques such as combining genes from various *Bt* strains. The transgenic engineering of *Bt* genes into cultivars is also a novel aspect of host plant resistance to insect pest attack. A principal difficulty is that of insect pest resistance which has already developed with *Bt* microbial insecticides. This resistance may be intensified by such broadscale pest exposure.

Table 9. Microbial Insecticides in Commercial Use or Development^a

Product	Pests controlled
<i>Bacteria</i>	
<i>Bacillus popilliae</i>	Japanese beetle larvae in soil
<i>Bacillus thuringiensis aizawa</i>	diamond back larvae, wax moth
<i>Bacillus thuringiensis israelensis</i>	mosquito and black fly larvae
<i>Bacillus thuringiensis kurstaki</i>	caterpillars
<i>Bacillus thuriengensis tenebrionis</i>	Colorado potato beetle
<i>Bacillus sphaericus</i>	mosquito larvae
<i>Fungi</i>	
<i>Beauveria bassiana</i>	European corn borer (China)
<i>Hirsutella thompsoni</i>	citrus red mite
<i>Metarhizium anisopliae</i>	sugar cane spittle bug (Brazil)
<i>Verticillium lecanii</i>	greenhouse aphids, whiteflies
<i>Protozoa</i>	
<i>Nosema locustae</i>	grasshoppers
<i>Viruses</i> ^b	
alfalfa looper NPV	alfalfa looper
beet armyworm NPV	beet armyworm
codling moth NPV	codling moth larvae (California, Europe)
gypsy moth NPV	gypsy moth larvae
heliiothis NPV	cotton bollworms, tobacco budworm
pine sawfly NPV	pine sawfly larvae
tussock moth NPV	Douglas fir tussock moth larvae

^a Ref. 51.
^b NPV = nuclear polyhedrosis virus.

Baculoviruses, especially nuclear polyhedrosis viruses (NPV) and granulosis viruses (GV), appear to be exceptionally well suited for IPM because of their extreme insect specificity. They are stomach poisons and are slow-acting. *In vitro* production is difficult and the products are more expensive than the bacterial insecticides. Their high host specificity is viewed as a commercial disadvantage, and improvements in formulations and application techniques are needed.

Fungi are broadly associated microbial pathogens of insects. *Beauveria bassiana* is extensively used as a microbial insecticide in China and Eastern Europe, especially for the control of the corn borer *Ostrinia* spp. However, fungi have had no significant development as microbial insecticides in the United States.

Microsporidia, especially *Nosema* spp., are pathogens in many insect pests. However, only *Nosema locustae*, which attacks grasshoppers, is marketed.

Insect Resistance to Insecticides

Many insect species have developed races that are sufficiently resistant to the action of specific insecticides so as to necessitate changes in control practices (52 ndash 57). This resistance, which has been described as accelerated microevolution, results from the selection of naturally occurring mutants that have acquired a resistant allele (R) that confers some degree of immunity to the insecticide through biochemical, physiological, or behavioristic factors. The R-allele may be dominant, recessive, or codominant and is spread through the pest population by preferential survival, following the intensive application of the insecticide. The development of resistance to dieldrin used in residual house spraying in West Africa to control the malaria vector *Anopheles gambiae*, is a classic example. R-alleles were found to be present in African mosquito populations at frequencies of 0.04 to 6.0^o so that a single residual spray treatment of dwelling produced an *Anopheles* population of 86^o RS and RR genotypes with virtual immunity to the insecticide.

History of Insecticide Resistance. Resistance to lime ndash sulfur spray was first described in 1914 in the San Jose scale *Quadraspidiotus perniciosus*. In 1916, the California red scale *Aonidiella aurantii* and the black scale *Saisettia oleae* attacking citrus were characterized as resistant to hydrogen cyanide, as applied by tent fumigation. The number of scientifically validated examples of resistance slowly increased and by 1946 with the advent of DDT, included the citricola scale *Coccus pseudomagnoliarum* to hydrogen cyanide, the codling moth *Cydia pomonella* and the peach twig borer *Anarsia lineatella* to lead arsenate, the citrus thrips *Scirtothrips citri* and the gladiolus thrips *Thrips simplex* to potassium antimonyl tartrate (tartar emetic), the walnut husk fly *Rhagoletis completa* to cryolite, and the cattle ticks *Boophilus decoloratus* and *B. annulatus* to sodium arsenite dip. The widespread use of DDT after World War II produced selection on a greatly increased scale and resistant races of the housefly *Musca domestica* appeared in 1946 in Sweden and Denmark, of the saltmarsh mosquito *Aedes sollicitans* in Florida, of the house mosquito *Culex pipiens* in Italy in 1947, of the bedbug *Cimex lectularius* in Hawaii in 1947, and of the human body louse *Pediculus humanus* in Korea and Japan in 1951 (52,54,55).

The successive commercialization of lindane and the cyclodienes, the organophosphates, the carbamates, and the pyrethroids resulted in the exponential development of insecticide-resistant species which totaled 69 by 1956, 224 by 1970, 364 by 1976, 428 by 1980, and 505 by 1988 (57). Two principal classifications of resistant species are encountered. Cross resistance enables resistant species to survive exposure to structurally related chemicals (eg, DDT resistance nonsusceptible to methoxychlor, lindane resistance nonsusceptible to dieldrin, and parathion resistance nonsusceptible to malathion). Cross resistance results from the action of a common detoxication system or a change in susceptibility of a common biochemical lesion, and has appeared chronologically to DDT methoxychlor, lindane cyclodienes, organophosphates, carbamates, and pyrethroids. Multiple resistance is far more serious and results from the common occurrence and interaction of survival mechanisms such as altered acetylcholinesterase producing multiple resistance to both organophosphates and carbamates, or nerve axon insensitivity that produces resistance to both DDT and pyrethroids. As a result there is an ever-widening pool of insect pests with multiple resistance which is found in at least 44 families of 10 insect orders. Once these genes are established in the genome, they have very lengthy persistence. Therefore, although removal of specific insecticidal pressures may decrease the frequency of the R gene, reapplication of the insecticide rapidly reselects for resistance. Genes for DDT and cyclodiene resistance in houseflies in Denmark have persisted since the 1970s and these insecticides again became ineffective after two months of application. The citrus thrip, *Scirtothrips citri*, has retained its resistance to tartar emetic since the late 1940s and to DDT since the late 1950s.

Insecticide resistance has had an enormous impact on insect control practices. In 1955, the World Health Organization inaugurated a global program to eradicate malaria by residual house spraying with DDT and dieldrin, to kill infective female *Anopheles* mosquito vectors before they could transmit the disease. Although DDT resistance was first observed in *Anopheles sacharovi* in Greece in 1950 and dieldrin resistance in 1954, the program was initially highly successful and malaria was eradicated in Sardinia and reduced to the vanishing point in Sri Lanka. However, resistance in important malaria vectors developed rapidly so that by 1986 there were a total of 66 resistant *Anopheles* species, 56 to DDT, 51 to dieldrin and lindane, 31 to organophosphates, 14 to propoxur, and 8 to permethrin. It became apparent that malaria could not be eradicated by the simple and inexpensive technology of residual house spraying.

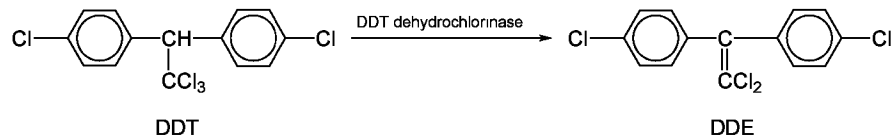
The corn rootworms *Diabrotica barberi* and *D. virgifera* were controlled in the U.S. corn belt from 1954 to 1964 by soil applications of heptachlor and aldrin. A resistant race of the western corn rootworm *D. virgifera* was first characterized in southeastern Nebraska in 1961 and spread rapidly throughout the entire corn belt. Successive introductions of carbamate and organophosphate soil insecticides have increased the cost of corn rootworm control from

about \$4.80 ha in 1960, to \$16.50 in 1975, and to as much as \$44 in 1990.

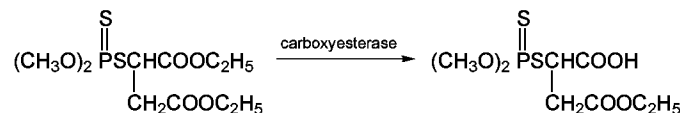
The hornfly *Haematobia irritans*, an important bloodsucking pest of cattle, was found to be completely controlled by ear tags which slowly released permethrin and fenvalerate. These were found to give almost 100% control for 24 weeks. However, following almost universal application of this ideal control measure, control failures of the pyrethroid insecticides were widely reported in North America in 1982–1983 and the practice is now abandoned.

Resistance Management. The lengthy persistence of resistant genes in insect genomes and the occurrence of multiple resistance are factors that have generally precluded the successful reuse of any insecticide to control insect populations with resistant alleles, even though the initial resistance has apparently reverted to susceptibility (55). Insecticide resistance is a complex genetic, ecologic, and evolutionary phenomenon and insecticide management practices using mixtures of insecticides, synergists, or alternative treatments with different classes of insecticides have not produced long-term solutions. Resistance management is most likely to be successful if it is aimed at reducing the frequency of the single-factored selection process resulting from conventional chemical control. IPM practices that reduce the frequency of exposure to specific insecticides, incorporate source reduction, and incorporate a variety of suppressive measures such as biological and cultural controls and host plant resistance offer the most practical principles for resistance management.

Biochemistry of Resistance. Studies of specific gene regulation of the physiological processes causing insecticide resistance have identified a number of increasingly generalized biochemical mechanisms. Metabolic resistance is a result of enhanced detoxication of the insecticide by enzymatic processes such as glutathion transferases, esterases, epoxide hydrolases, and microsomal oxidases. Examples include DDTase, a glutathion dependent enzyme that attacks the α -hydrogen of DDT resulting in the expulsion of Cl to form the noninsecticidal DDE.



This resistance mechanism has been identified in the housefly, a variety of mosquito species, *Triatoma infestans*, the human louse, the pink bollworm *Pectinophora gossypiella*, the red-banded leafroller *Argyrotaenia velutinana*, and the Mexican bean beetle *Epilachna varivestis*. DDTase is inactivated by DDT analogues such as the acaricide chlorfenethol or 1,1-bis-(4-chlorophenyl)-ethanol [80-06-8], and by 4-chlorobenzene N,N-dibutylamide [127-59-3]; these have had limited use as synergists for resistant houseflies. Malathion resistance in the mosquito *Culex tarsalis* and in the green rice leafhopper *Nephotettix viaticipes* results from enhanced carboxyesterase enzyme that attacks one of the carboethoxy groups to form a noninsecticidal compound.



Carboxyesterase is inhibited by the organophosphorus insecticide EPN in its activated oxon form and this has been used as a malathion synergist.

Knockdown resistance (kdr) results from selection for reduced sensitivity of the sodium channels of the nerve axon that are the target site for DDT and pyrethroids. Kdr resistance is common in the housefly, hornfly, mosquito, bedbug, and many chewing caterpillars. Altered acetylcholinesterase is a biochemical change in the structure of this enzyme essential for nerve impulse transmission, that retards inhibition by organophosphorus and carbamate insecticides. It has been characterized in the housefly, the green rice leafhopper, and in red spider mites. Penetration (pen) resistance slows penetration of the insecticides to the site of action. A variety of these resistance mechanisms may be present in a single insect pest species obviating the effective use of synergists, and they may act in concert to increase resistance levels. From such multiple resistance insect pest species may obtain virtual immunity to large groups of insecticides such as pyrethroids, organophosphates, and carbamates.

Insecticide Formulation

The successful employment of any insecticide depends on its proper formulation into a preparation that can be applied for insect control with safety to the applicator, animals, and plants. Insecticides are commonly formulated as dusts, water dispersions, emulsions, and solutions. The preparation and use of such formulations involves accessory agents such as dust carriers, solvents, emulsifiers, wetting and dispersing agents, stickers, and deodorants or masking agents (1).

Dusts are the simplest formulations and generally contain low concentrations, 0.1 – 20 %, of the toxicant, although ground botanical preparations may be used undiluted. Therefore, the properties of the carrier largely determine the quality of the finished dust. Carriers commonly include organic flours, sulfur, silicon oxides, lime, gypsum, talc, pyrophyllite, bentonites, kaolins, attapulgite, and volcanic ash. Selection of the carrier is made on the basis of compatibility with the desired insecticide (including pH, moisture content, and stability), particle size, abrasiveness, absorbability, density, wettability, and cost. The mixture of the toxicant and diluent is made by a variety of simple operations, eg, milling, solvent impregnations, fusing, and grinding. The particle size usually ranges from 0.5 – 4.0 μm in diameter.

Wettable powders are prepared by blending the toxicant in high concentration, usually from 15 to 95%, with a dust carrier such as attapulgite which wets and suspends properly in water. One to two percent of a surface-active agent usually is added to improve the wetting and suspensibility of the powder. Sprays of wettable powders are used widely in agriculture because of their relative safety to plants.

Granulars are pelleted mixtures of toxicant, usually at 2.5–10%, and a dust carrier, eg, absorptive clay, bentonite, or diatomaceous earth, and commonly are 250–590 µm in particle size. They are prepared by impregnation of the carrier with a solution or slurry of the toxicant and are used principally for mosquito larviciding and soil applications.

Emulsives are solutions of toxicant in water-immiscible organic solvents, commonly at 15–50%, with a few percent of surface-active agent to promote emulsification, wetting, and spreading. The choice of solvent is predicated upon solvency, safety to plants and animals, volatility, flammability, compatibility, odor, and cost. The most commonly used solvents are kerosene, xylenes and related petroleum fractions, methyl isobutyl ketone, and amyl acetate. Water emulsion sprays from such emulsive concentrates are widely used in plant protection and for household insect control.

Baits include mixtures of toxicant, usually at 1–5%, with a carrier especially attractive to the insect pest. Carriers include sugar for the houseflies, protein hydrolysates for fruit flies, bran for grasshoppers, and honey, chocolate, or peanut butter for ants.

Slow release formulations incorporate nonpersistent compounds, eg, methyl parathion, insect growth regulators, and sex pheromones, in a variety of granular, laminated, microencapsulated, and hollow-fiber preparations.

Application. The usefulness of any insecticide is substantially dependent upon its proper application and this is determined by the properties of the insecticide, the habits of the pest to be controlled, and the site of the application to be made. The three general methods of applying insecticides are spraying, with water or oil as the principal carrier; dusting, with a fine dry powder as the carrier; and fumigation, where the insecticide is applied as a gas (1).

The proper choice and application of an insecticide for pest control are predicated upon factors, eg, the life history and ecology of the pest, the relation of pest population to economic damage, the effect of the insecticide on the pest or its plant or animal host, related organisms in the ecosystem, and proper timing of the application to prevent illegal residues at harvest and to avoid damaging of bees and other pollinating insects.

Sprays are the most common means of insecticide application and generally involve the use of water as the principal carrier, although volatile oils sometimes are used. With the older inorganic insecticides, suspensions in water were used at dilutions of 0.1 to 0.2%. The development of the more effective organic insecticides has allowed the widespread use of concentrate sprays in which the toxicant is contained at 10 to 98% and the amount of carrier to be applied is enormously reduced. The use of concentrate or ultralow volume sprays has brought about a revolution in spray equipment away

from hydraulic nozzles with coarse atomization producing droplets of 200 – 500 μm in diameter to air blast and other atomizing nozzles producing droplets of 30 – 80 μm in diameter. The use of concentrate sprays also has assisted the development of spraying by airplane and helicopter.

Aerosols (qv) are very finely divided sprays having droplet diameters of 1 – 30 μm. They are used almost entirely as space sprays for application to enclosures, particularly against flying insects. Aerosols are most conveniently applied by the familiar liquefied gas dispersion or bomb but can be generated on a larger scale by rotary atomizers or twin fluid atomizers.

Dusts are the simplest means of insecticide dispersal and are applied by introducing the finely divided carrier, with particles of 0.5 – 3.0 μm in diameter, into a moving air stream. In comparison with sprays, dusts adhere poorly to surfaces and cause serious drift problems away from the treatment area.

Genetic Control. Manipulation of the mechanisms of inheritance of the insect pest populations has occurred most successfully through the mass release of sterilized males, but a variety of other techniques have been studied, including the environmental use of chemosterilants and the mass introduction of deleterious mutations, eg, conditional lethals and chromosomal translocations (58 – 60) (see GENETIC ENGINEERING).

Sterile Male Release. The release of large numbers of sterile male insects into a population of virgin females has been shown (58) to reduce the number of fertile females in subsequent generations, as shown in Table 10. The most practical utilization of the sterile male technique has been in the eradication of the screwworm fly *Cochliomyia hominivorax* from the southeastern United States in 1959, which involved the release of ca2 × 10⁹ male flies, sterilized with gamma radiation from ⁶⁰Co, over an area of 181,000 km² for 18 months. The eradication program was extended to the Texas–Mexico border and has successfully controlled the screwworm for over more than 25 years through annual release of as many as one billion sterile males. The screwworm was established in Libya in 1988 and subsequently became widely distributed through North Africa. Application of the sterile male release program has apparently eradicated this pest from Libya and may have prevented further infestation of the continent.

Table 10. Theoretical Effects of Release of Sterile Male^a Insects on a Natural Insect Population^b

Generation	Wild female population	Wild females mated to sterile males, %	Theoretical population of fertile females
0th	1,000,000	66.7	333,333
1st	333,333	85.7	47,619
2nd	47,619	97.7	1107
3rd	1107	99.95	1

^a Population released = 2,000,000.

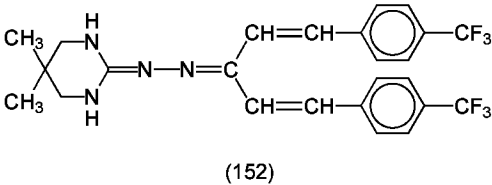
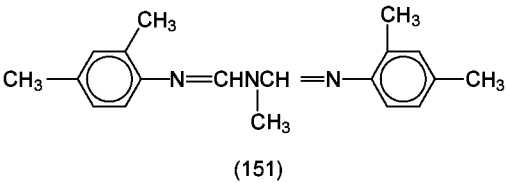
^b Ref. 58.

Mass releases of sterile male insects have produced dramatic reductions in the populations of the Mediteranean fruit fly *Ceratitis capitata* in California beginning in 1981 when 40 million sterile flies were released weekly; and in the codling moth *Cydia pomonella* in isolated apple orchards in the Pacific Northwest.

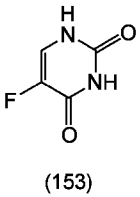
The prerequisites for successful control programs involving the release of sterile male insects depend on (1) a method for mass rearing of the insect pest, (2) adequate dispersion of the released sterile males, (3) a sterilization procedure that does not adversely affect mating behavior, (4) a pest species where mating occurs only once per lifetime or one in which the sperm from sterile males competes with those from fertile males, and (5) a low initial population density or a means of reduction of the population to levels that permit the release of a dominant population of sterile males. However, many species of pest insects do not conform to these criteria and the release of sterile males is an expensive procedure, best suited to large-scale governmental programs.

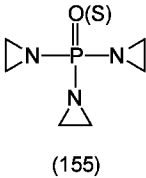
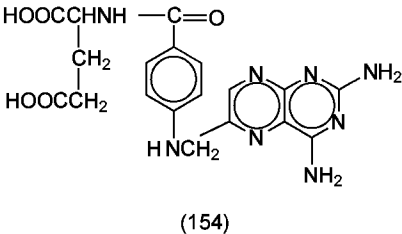
Chemosterilants. The use of chemicals that would sterilize large segments of natural insect pest populations has been suggested (61). Several types of chemosterilants are known to produce adequate sterility in insects by preventing the production of ova or sperm, by causing death of sperm or ova, or by producing severe injury to the genetic material of sperm or ova so that the zygotes that are produced do not develop into mature progeny.

Antimetabolites include 5-fluorouracil [51-21-8] (153) and amethopterin [59-05-2] (154), a folic acid antagonist, which produce sterility in female flies when fed at 0.01–0.05% in the diet.



Radiomimetic compounds include cancer chemotherapeutic compounds that incorporate the extremely reactive ethyleneimine group. TEPA [57-39-6] or 1-tris(1-aziridinyl)phosphine oxide (155) (mp 41°C, rat oral LD₅₀ 37 mg/kg) and its sulfur analogue, thiotepa [52-24-4] (mp 51.5°C) were among the first to be investigated. These compounds are alkylating agents for DNA and cause sterility in both sexes of the housefly, eg, when incorporated into food or applied topically at 0.1–1%. Mosquitoes and flies are sterilized by exposure to surface residues of 110–2690 mg/m². Other similar radiomimetic compounds include apholate [52-46-0] or 2,2,4,4,6,6-hexa-(1-aziridinyl)-2,4,6-triphospho-1,3,5-triazine (156) (mp 155°C, rat oral LD₅₀ 90 mg/kg), and hempa [680-31-9] or hexamethyltriphosphoramide (157) (rat oral LD₅₀ 2600 mg/kg). Busulfan [55-98-1], 1,4-butanediol dimethanesulfonate, CH₃SO₂O(CH₂)₄OSO₂CH₃ (mp 114–118°C), has a rat LD₅₀ of 1.8 mg/kg (intravenous) and has been used for the mass production of sterile insects.





Unfortunately, all of these compounds, including hempa, are mutagens and strong carcinogens and the radiomimetic compounds produce degenerative changes in human germ plasm at dosages of 0.01 × LD₅₀ or less. Their casual and indiscriminate use in insect control, therefore, is precluded. Successful application of the chemosterilant principle must await the availability of more selective insect sterilants that do not specifically affect the universal DNA–RNA genetic mechanism.

Repellents and Attractants.

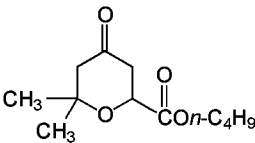
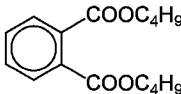
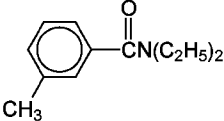
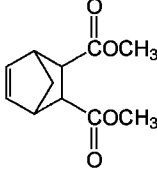
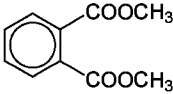
Repellants. Repellents (qv) are substances that protect animals, plants, or products from insect attack by making food or living conditions unattractive or offensive (62–65). These substances, which may not be poisonous or only mildly toxic, are rarely, if ever, effective against all kinds of insects. Such chemicals can sometimes be employed to advantage where it is impossible to use an insecticide and may afford a greater or lesser degree of protection to manufactured products, growing plants, or the bodies of animals and humans. Among the many examples are the following. (1) Repellents against crawling insects. Examples are the creosote lines used as barriers to the migration of chinch bugs; trichlorobenzene and other chemicals used to protect buildings from termites; heavy oils at the base of poultry roosts as a barrier to poultry mites; and certain chemical bands about tree trunks. (2) Repellents against the feeding of insects. These include the application of bordeaux, lime, and similar washes to plants to ward off leafhoppers and some chewing insects; mosquito repellents and fly sprays to lessen the attacks of bloodsucking flies and mosquitoes; the application of sulfur to the body to keep chiggers from attacking; the use of smoke and smudges to repel biting flies; the chemical treatment of logs to keep beetle borers from destroying log cabins and other rustic work; and moth balls, oil of cedar, and mothproofing treatments to protect materials from attack by clothes moths and carpet beetles. (3) Repellents against the egg laying of insects. Examples are the use of pine-tar oil and diphenylamine to keep screwworm flies from laying eggs about wounds of animals.

Bordeaux Mixture. Bordeaux mixture (see FUNGICIDES) originated in France as a spray to control the downy mildew disease (caused by the fungus *Peronospora viticola*) of grapes in ca 1882, and was first used in the United States in 1887. Although primarily a fungicide, bordeaux mixture is repellent to many insects, eg, flea beetles, leafhoppers, and potato psyllid, when sprayed over the leaves of plants. It is, to some extent, an ovicide and has some residual toxic effect upon the sap and thereby kills leafhoppers and psyllids for some days. Bordeaux mixture is produced by mixing hydrated lime, 3.6–4.5 kg (8–10 lb), and copper sulfate, 1.9–2.7 kg (4–6 lb), in 380 L (100 gal) of cold water to produce a precipitate of tetracupric sulfate, 4CuO ·SO₃, and pentacupric sulfate 5CuO ·SO₃. The colloid remains in suspension for several hours and when sprayed on foliage covers the leaves with a thin, tenacious film which gradually produces soluble copper.

Repellents to Bloodsucking Insects. The limiting criteria for a good repellent against bloodsucking insects are, in order of importance, effective protection of the treated area for several hours on all types of subjects and under all climatic conditions; complete freedom from toxicity and irritation when regularly applied to human or animal skin; cosmetic acceptability, including freedom from unpleasant odor, taste, and touch, and harmlessness to clothing; protection against a wide variety of biting insects; and low cost and availability (63). Tests of many thousands of chemical compounds for repellent action to flies, mosquitoes, chiggers, fleas, and ticks have shown that, although many possess a significant degree of repellency to the various pests, few meet all the other requirements satisfactorily. The materials that have found practical application as repellents are listed in Table 11 with their properties and uses. Those most widely used by out-of-doors persons are *N,N*-diethyl *m*-toluamide (DEET), dimethyl phthalate, and 2-ethyl-1,3-hexane diol (alternative name "6-12").

Table 11. Repellents for Mosquitoes, Flies, Mites, and Ticks^a

Name	CAS Registry Number	Formula	Properties	Toxicity, oral LD ₅₀ rat, mg/kg	Uses
benzil	[134-81-6]		solid, mp 95°C		clothing treatment for chiggers
benzyl benzoate	[120-51-4]		oily liquid, bp 323°C, mp 21°C, <i>d</i> 1.12 g/cm ³	1,900	clothing treatment for chiggers
butoxypoly-(propylene glycol)	[9003-13-8]	$n\text{-C}_4\text{H}_9\text{O}-(\text{CH}_2\underset{\text{CH}_3}{\text{CHO}})_n\text{-CH}_2\underset{\text{CH}_3}{\text{CHOH}}$	liquid, 400 and 800 mol wt fractions, <i>d</i> 0.973–0.990 g/cm ³	11,200	fly repellent for cattle
<i>N</i> -butyl-acetanilide	[91-49-6]		liquid, bp 277–281°C, mp 22°C, <i>d</i> 0.99 g/cm ³	2,830	clothing treatment for fleas, ticks
indalone	[532-34-3]		brown liquid, bp	7,800	mosquito and fly

			110–115°C 0.33 kPa, ^b 1.06 g cm ³		repellent
dibutyl adipate	[105-99-7]	$n\text{-C}_4\text{H}_9\text{OOCCH}_2\text{CH}_2\text{CH}_2\text{CH}_2\text{COO}n\text{-C}_4\text{H}_9$ H ₉	liquid, bp 193°C 1.9 kPa, ^b d 0.965 g cm ³	12,900	tick repellent
dibutyl phthalate	[84-74-2]		liquid, bp 340°C, d 1.045 g cm ³	21,000	clothing treatments for chiggers
di- <i>n</i> -butyl succinate	[141-03-7]	$n\text{-C}_4\text{H}_9\text{OOCCH}_2\text{CH}_2\text{COO}n\text{-C}_4\text{H}_9$	liquid, bp 108°C 0.53 kPa, ^b mp 29°C	8,000	fly repellent for cattle
2-hydroxyethyl- <i>n</i> -octyl sulfide	[3547-33-9]	CH ₃ (CH ₂) ₇ SC ₂ H ₄ OH	liquid, bp 98°C at 13 Pa mp 0°C, d 0.93 g cm ³	8,500	space repellent
<i>N,N</i> -diethyl- <i>m</i> -toluamide (deet)	[134-62-3]		liquid, bp 111°C 0.133 kPa, ^a d 0.996 g cm ³	2,000	general-purpose repellent
dimethyl carbate	[39589-98-5]		solid, mp 40°C	1,000	mosquito repellent
dimethyl phthalate	[131-11-3]		liquid, bp 284°C, d^{25} 1.189 g cm ³	8,200	general-purpose repellent
2-ethyl-2-butyl-1,3-propanediol	[115-84-4]	$\text{CH}_3\text{CH}_2\text{C}(\text{CH}_2\text{OH})_2\text{C}_4\text{H}_9$	solid, mp 40°C, d 0.931 g cm ³	5,040	mosquito repellent
2-ethyl-1,3-hexanediol (6-12)	[94-96-2]	$\text{HOCH}_2\text{CH}(\text{C}_2\text{H}_5)\text{CH}(\text{OH})\text{CH}_2\text{CH}_2\text{CH}_3$	liquid, bp 244°C, d 0.942 g cm ³	2,400	general-purpose repellent
di- <i>n</i> -propyl isocinchonate	[136-45-8]	$n\text{-C}_3\text{H}_7\text{OOC}$  $\text{COO}n\text{-C}_3\text{H}_7$	liquid, bp 186°C 0.133 kPa, ^b d 1.08 g cm ³	6,230	fly repellent for cattle
<i>n</i> -propyl <i>N,N</i> -diethylsuccinamate	[5834-44-4]	$\text{C}_3\text{H}_7\text{OCCH}_2\text{CH}_2\text{CNC}(\text{C}_2\text{H}_5)_2$	d 1.01 g cm ³	6,400	mosquito repellent

^a Ref. 63 and 64.
^b To convert kPa to mm Hg, multiply by 7.5.

Repellent compounds exhibit wide differences in their activity against various species of mosquitoes and flies, as well as other biting arthropods, and it has been demonstrated that the overall protection against various pests is greatly extended by the use of mixtures, eg, dimethyl phthalate:indalone:2-ethyl-1,3-hexanediol, 3:1:1 parts by wt; and dimethyl phthalate:2-ethyl-1,3-hexanediol:dimethyl carbate, 4:3:3 parts by wt. These repellents can be incorporated into various creams and lotions. Applications of the most effective materials give from one to six hours protection against mosquitoes and biting flies.

Under many circumstances, the ideal applications of repellents are those made on clothing, gloves, and head nets where the protection time is extended to a week or more. Clothing applications may be made by rubbing or spraying the repellent on the cloth or by dipping the clothing in an emulsion of the repellent. The clothing-impregnation method is especially suited to military operations; the formulation for the standard clothing impregnant, M-1960, of the U.S. Armed Services for protection from mosquitoes, fleas, ticks, and chiggers is benzyl benzoate:*N*-butylacetanilide:2-ethyl-2-butyl-1,3-propanediol:Tween 80, 3:3:3:1.

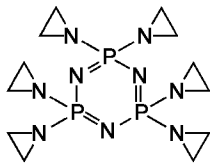
The pyrethroid permethrin is an effective repellent for biting flies, mosquitoes, and ticks. A 30% permethrin cream is a general-purpose repellent used by the U.S. Armed Forces.

Mothproofing. Rugs, carpets, upholstered furniture and woolen clothes are damaged to an estimated \$1 to \$2 billion annually in the United States by fabric pests (66). There are two main groups of insects that can digest the keratinaceous proteins of wool and of animal hides and utilize them as complete dietary sources. The beetles of the family Dermestidae include as principal pests the carpet beetle, *Anthrenus scrophulariae*, the furniture carpet beetle, *A. flavipes*, the varied carpet beetle, *A. verbasci*, the black carpet beetle, *Attagenus unicolor*, and the wardrobe beetle, *A. fasciatus*. The larvae of the clothes moths of the family Tineidae have similar feeding habits. The important species are the casemaking clothes moth, *Tinea pellionella*, the webbing clothes moth, *Tineola bisselliella*,

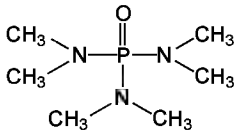
and the carpet moth, *Trichophaga tapetælla*. Mothproofing materials to protect against damage are of two general classes: those that can be applied by the householder to give more or less temporary protection; and colorless dyestuffs fixed through chemical reaction with wool protein that are applied during the manufacture of rugs and upholstery.

The persistent pyrethroids such as permethrin, cypermethrin, cyfluthrin, and fenvalerate are effective mothproofing agents when applied as spot treatments from ready to use (RTU) formulations. Sodium fluosilicate [16893-85-9] is an effective mothproofing agent used at 0.5 to 0.7% in water solution with 0.3% potassium aluminate and 0.03% oxalic acid, and applied to fabrics by spraying or dipping. It is not removed by dry cleaning.

Colorless dyes of the chlorinated diphenyl ether–ureasulfonic acid type (Mitin [3567-25-7] (158)) and the (polychloro-2-chloromethylsulfonamido)-diphenylether type (Eulan (159)) applied at 1–3% of the fabric weight are the most widely used during the fabrication of woolens. They cannot be removed by dry cleaning and provide protection from fabric pests over the lifetime of the product.



(156)



(157)

DDT at 0.25–0.75% and dieldrin at 0.05% applied during dry cleaning or by spraying fabrics with aqueous emulsions of equivalent strength were formerly used extensively as semipermanent fabric protectants. However, these are no longer used because of problems of environmental contamination.

Attractants. The use of specific insect attractants is an important component of modern IPM technology (67–71) and promotes essentially nonpolluting and more economical pest control practices. These chemicals may have attractant, arrestant, and phagostimulant properties. They are used (1) to monitor insect populations in relation to the economic threshold, and to time control operations using a variety of inexpensive sticky traps of flat, cylindrical, and delta design; (2) to detect incipient infestations of exotic pests; (3) for removal trapping as with Japanese beetle, bark beetles, and tsetse flies; (4) to lure insects to toxic baits; and (5) with certain sex pheromones applied at relatively high dosages, such as for the gypsy moth, codling moth, pink bollworm, and grape berry moth to cause mating disruption. The semichemicals used are categorized as (1) sex pheromones involved in intraspecific chemical communication; (2) aggregation pheromones promoting mass attack on host plants; and (3) plant kairomones involved in interspecific chemical communication leading to host plant selection. Attractants for more than 250 species of insect pests are available commercially (70).

Sex Pheromones. These have been identified from several hundred insect species of 10 orders (68,70). Their evolutionary development is most advanced in the Lepidoptera where they are typically produced by eversible glands opening at the tip of the female abdomen. In this order they are long-chain unsaturated alcohols, aldehydes, and esters ranging from C₈ to C₁, that are perceived by males in picogram to nanogram quantities and can lure them from distances of 10 to 1000 m or more. Specific sex pheromone chemicals used in important IPM programs are listed in Table 12. The naturally occurring sex pheromones are often blends, and maximum responses to synthetic mixtures depend on specific chemical composition. For example, the male European corn borer, *Ostrinia nubilalis*, is attracted by a mixture of the (Z)- and (E)-isomers of 11-tetradecenol acetate. The midwestern U.S. race is optimally attracted by a 97:3 mixture of these isomers and the eastern race by a 3:97 mixture. The male corn earworm (bollworm) *Heliothis æa* is optimally attracted by a blend of (Z)-7, (Z)-9, and (Z)-11-hexadecenal, and hexadecanol.

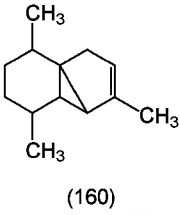
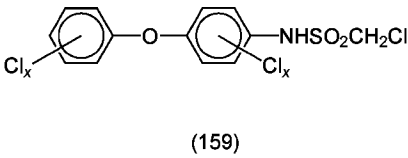
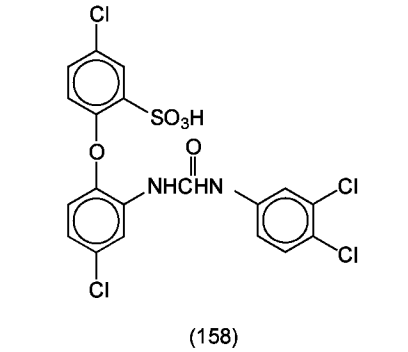
Table 12. Sex Pheromones Used in Insect Control^a

Compound	CAS Registry Number	Structure	Species
(Z)-7,8-epoxy-2-methyl-octade cane	[29804-22-6]	$\text{CH}_3(\text{CH}_2)_9\text{CH} \begin{array}{c} \diagup \diagdown \\ \text{O} \end{array} \text{CH}(\text{CH}_2)_4\text{CH}(\text{CH}_3)_2$	gypsy moth, <i>Porthetria dispar</i>
(E)-8,(E)-10-dodecadienol	[33956-49-9]	$\text{CH}_3\text{CH}=\text{CHCH}=\text{CH}(\text{CH}_2)_2\text{CH}_2\text{OH}$	codling moth, <i>Cydia pomonella</i>
(Z)-9-tetradecenal (with (Z)-11-hexadecenal)	[53939-27-8]	$\text{CH}_3\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}=\text{CH}(\text{CH}_2)_7\text{CHO}$	tobacco budworm, <i>Heliothis virescens</i>
(E)-11-tetradecenal	[35746-21-5]	$\text{CH}_3\text{CH}_2\text{CH}=\text{CH}(\text{CH}_2)_9\text{CHO}$	eastern spruce budworm, <i>Choristoneura fumiferana</i>
(Z)-11-hexadecenal	[53939-28-9]	$\text{CH}_3(\text{CH}_2)_3\text{CH}=\text{CH}(\text{CH}_2)_9\text{CHO}$	corn earworm, <i>Heliothis æa</i>
(Z,Z)-11,13-hexadecadienal	[71317-73-2]	$\text{CH}_3\text{CH}_2\text{CH}=\text{CHCH}=\text{CH}(\text{CH}_2)_9\text{CHO}$	navel orangeworm, <i>Amyelois transitella</i>
(E)-5-decenyl acetate		$\text{CH}_3(\text{CH}_2)_3\text{CH}=\text{CH}(\text{CH}_2)_4\text{OOCCH}_3$	peach twig borer, <i>Anarsia lineatella</i> (with (E)-5-decenol)
(Z)-7-dodecenyl acetate	[14959-86-5]	$\text{CH}_3(\text{CH}_2)_3\text{CH}=\text{CH}(\text{CH}_2)_2\text{OOCCH}_3$	cabbage looper, <i>Trichoplusia ni</i> , alfalfa looper, <i>Autographa californica</i> , soybean looper, <i>Pseudoplusia includens</i> , black cutworm, <i>Agrotis ipsilon</i> (with (Z)-9-14 Ac,(Z)-11-16Ac)
(E,Z)-10,12-hexadecadienal		$\text{CH}_3\text{CH}_2\text{CH}_2\text{CH}=\text{CHCH}=\text{CH}(\text{CH}_2)_8\text{CHO}$	tobacco hornworm, <i>Manduca sexta</i>
(Z)-8-dodecenyl acetate	[28079-04-1]	$\text{CH}_3(\text{CH}_2)_2\text{CH}=\text{CH}(\text{CH}_2)_7\text{OOCCH}_3$	oriental fruitmoth, <i>Grapholitha molesta</i>
(Z)-9-dodecenyl acetate	[16974-11-1]	$\text{CH}_3\text{CH}_2\text{CH}=\text{CH}(\text{CH}_2)_8\text{OOCCH}_3$	grape berry moth, <i>Paralobesia viteana</i> (with (E)-isomer)
(E)-10-dodecenyl acetate		$\text{CH}_3\text{CH}=\text{CH}(\text{CH}_2)_9\text{OOCCH}_3$	tentiform leafminer, <i>Phyllonorycter blancardiella</i>
(Z)-9-tetradecenyl acetate	[16725-53-4]	$\text{CH}_3(\text{CH}_2)_3\text{CH}=\text{CH}(\text{CH}_2)_8\text{OOCCH}_3$	southern armyworm, <i>Spodoptera eridania</i> (with (Z)-9,(E)-12-tetradecadienyl acetate) beet armyworm, <i>Spodoptera exigua</i>

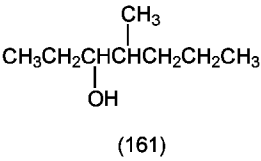
(Z)-11-tetradecenyl acetate	[20711-10-8]	$\text{CH}_3\text{CH}_2\text{CH}=\text{CH}(\text{CH}_2)_{10}\text{OOCCH}_3$	fall armyworm, <i>Spodoptera frugiperda</i> (with (Z)-9,(E)-12-tetradecadienyl acetate) red-banded leafroller, <i>Argyrotaenia velutinana</i> European corn borer, <i>Ostrinia nubilalis</i>
(E)-11-tetradecenyl acetate (with (Z)-11)	[39298-60-7]	$\text{CH}_3\text{CH}_2\text{CH}=\text{CH}(\text{CH}_2)_{10}\text{OOCCH}_3$	
(Z)-9,(E)-11-tetradecadienyl acetate (with (Z)-9,(E)-12)	[30562-09-5]	$\text{CH}_3\text{CH}_2\text{CH}=\text{CHCH}=\text{CH}(\text{CH}_2)_8\text{OOCCH}_3$	cotton leafworm, <i>Spodoptera littoralis</i> , <i>Spodoptera litura</i>
(Z)-9,(E)-12-tetradecadienyl acetate	[31654-77-0]	$\text{CH}_3\text{CH}=\text{CHCH}_2\text{CH}=\text{CH}(\text{CH}_2)_8\text{OOCCH}_3$	almond moth, <i>Cadra cautell</i> fig moth, <i>Cadra figulilella</i> Mediterranean flour moth, <i>Anagasta kuehniella</i> tobacco moth, <i>Ephestia elutella</i> pink bollworm, <i>Pectinophora gossypiella</i>
(Z)-7,(Z)-11-hexadecadienyl acetate (with (Z)-7,(E)-11)	[52287-99-5]	$\text{CH}_3(\text{CH}_2)_3\text{CH}=\text{CH}(\text{CH}_2)_2\text{CH}=\text{CH}(\text{CH}_2)_2\text{OOCCH}_3$	Angoumois grain moth, <i>Sitotroga cerealella</i> peach-tree borer, <i>Synanthedon exitiosa</i>
(Z)-3,(Z)-13-octadecadienyl acetate (with (E)-3,(Z)-13)	[53120-27-7]	$\text{CH}_3(\text{CH}_2)_3\text{CH}=\text{CH}(\text{CH}_2)_8\text{CH}=\text{CH}(\text{CH}_2)_2\text{OOCCH}_3$	
(E)-3,(Z)-13-octadecadienyl acetate	[53120-26-6]	$\text{CH}_3(\text{CH}_2)_3\text{CH}=\text{CH}(\text{CH}_2)_8\text{CH}=\text{CH}(\text{CH}_2)_2\text{OOCCH}_3$	lesser peach-tree borer, <i>Synanthedon pictipes</i>

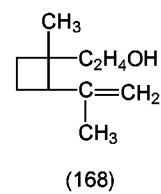
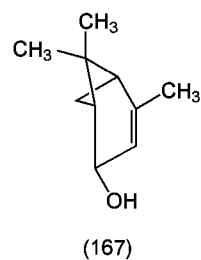
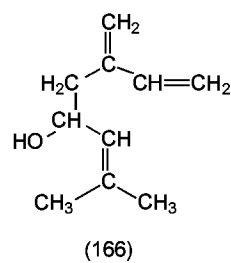
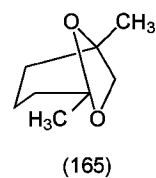
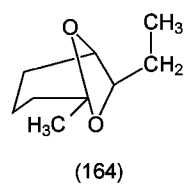
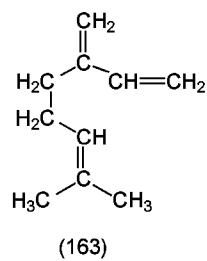
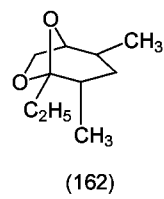
^a Ref. 70.

Aggregation Pheromones. Bark beetles of the family Scolytidae especially those of the genera *Dendroctonus*, *Ips*, *Phloeosinus*, and *Scolytus* are devastating enemies of fruit, shade, and forest trees (67,68,70). The depredations of the lesser European bark beetle, *Scolytus multistriatus*, in spreading the pathogen of the Dutch elm disease caused by the fungus *Ceratocystis ulmi*, are well known. Mass attacks by this species are the result of a complex aggregation pheromone composed of the terpenoid α -cubebene [17699-14-8] (**160**) produced by dying elms, which stimulates the production of the beetle frass pheromones 4-methyl-3-heptanol [14979-39-6] (**161**) and α -multistriatin,(1S)-endo,endo-5-ethyl-2,4-dimethyl-6,8-dioxabicyclo-[3.2.1]-octane [54832-20-1] (**162**). This pheromone complex is available as multilure.



The western pine beetle *Dendroctonus brevicomis* is perhaps the most destructive insect enemy of western pine forests. The aggregation pheromone is a mixture of the terpenoid myrcene [123-35-3] (**163**) from the tree and the frass pheromones *exo*-brevicommin [20290-99-7] (**164**) and frontalin [28401-39-0] (**165**). The Norway spruce beetle *Ips typographus* converts the tree terpenoid myrcene into the frass pheromone ipsdienol [35628-00-3] (**166**) and the beetles also produce 2-methyl-3-buten-2-ol [115-18-4], and *cis*-verbenol [473-67-6] (**167**), all of which are components of the aggregation pheromone.

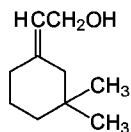




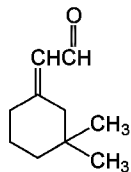
Similar pheromone blends are produced by other species of bark beetles and these complex aggregation pheromones have been used extensively in mass trapping programs, capturing millions of bark beetles and preventing them from further attacking valuable forest resources.

The aggregation pheromone of the boll weevil, *Anthonomus grandis*, is a mixture of the alcohols *D-cis*-2-isopropenyl-1-methylcyclobutaneethanol [30820-22-5] (168) and *cis*-3,3-dimethyl- Δ^1 , β -cyclohexaneethanol [26532-23-0] (169) and of the *cis*- and *trans*-isomers of the aldehyde of the latter (170). The pheromone is marketed as grandlure [11104-05-5] for monitoring and removal trapping of boll weevil populations.

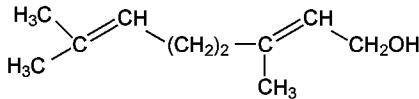
Kairomone Attractants. Semiochemicals synthesized by plants are used by many insect pests as cues for host plant selection, as attractants, arrestants, and feeding and oviposition stimulants (71). They may be attractive in nanogram to microgram quantities and are used for monitoring insect infestations, for removal trapping, for timing control operations, and as lures in poison baits. The best known kairomone attractants or food lures are geraniol [106-24-1] (171), eugenol [97-53-0] (172), and 2-phenylethanol propionate [122-70-3] (173) which are used in a mixture (3.5:3.5:3) as the attractant (japonilure) for the adult Japanese beetle, *Popillia japonica*, employed in hundreds of thousands of bag traps to protect gardens and orchards from this voracious pest.



(169)

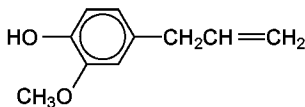


(170)

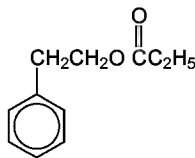


(171)

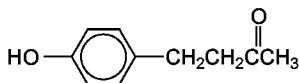
Methyl eugenol or 3,4-dimethoxy-1-allylbenzene [95-15-2] was first characterized in 1915 as a powerful attractant for the male oriental fruit fly, *Dacus dorsalis*, and attracts at least 60 other closely related *Dacus* spp. Raspberry ketone [5471-51-2] or 1-(4-*p*-hydroxyphenyl)-2-butanone [5471-51-2] (**174**) is an equally powerful attractant for the melon fly, *Dacus cucurbitae*, and the Queensland fruit fly, *D. tryoni*, and at least 180 other closely related *Dacus* spp. The acetyl ester (cue-lure) (**175**) is more volatile and is a synthetic parakairomone especially effective for monitoring infestations by these species. Methyl eugenol and cue-lure [3572-06-3] have been used successfully in "male annihilation" of the oriental fruit fly and the melon fly by applying them to fiber board blocks or pieces of twine together with malathion or naled insecticides and distributing them over infested areas at doses of 15 g of attractant and 1 g of the insecticide per ha.



(172)



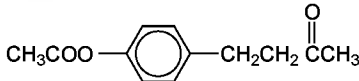
(173)



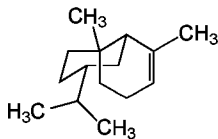
(174)

The male Mediterranean fruit fly *Ceratitidis capitata* is similarly attracted to the terpenoid α -copaene [3856-25-5] (**176**) from the oil of *Angelica archangelica* and this and the parakairomone *tert*-butyl 2-methyl-4-chlorocyclohexanoate (trimedlure [12002-53-8]) are very extensively employed in monitoring for infestations of this destructive pest. The female apple maggot fly *Rhagoletis pomonella* is attracted to the apple volatile butyl hexanoate, which is used to bait sticky red spheres to monitor populations and time spray treatments.

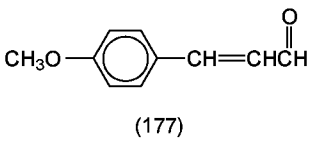
The *Diabrotica* spp. corn rootworm beetles are specifically attracted to a variety of plant-produced phenylpropanoids, eg, (*E*)-cinnamaldehyde [14371-10-9] for the southern corn rootworm *D. undecimpunctata howardi*; (*E*)-cinnamyl alcohol [4407-36-7] for the northern corn rootworm *D. barberi*; and indole [120-72-9] for the western corn rootworm, *D. virgifera virgifera*. Especially powerful lures for these rootworm beetles are 2-(4-methoxyphenyl)ethanol for the northern corn rootworm and 4-methoxycinnamaldehyde [71277-11-7] (**177**) for the western corn bootworm.



(175)



(176)



These lures are used with sticky traps to monitor infestations and in toxic baits together with the triterpenoid feeding stimulants, cucurbitacins B and E. Other lures for Coleoptera include caproic acid [142-62-7] for the green June beetle, *Cotinus nitida*, propyl 1,4-benzodioxane-2-carboxylate [24902-02-1] (178) for the European chafer, *Rhizotrogus majalis*, and ethyl 3-isobutyl-2,2-dimethylcyclopropanoate [33419-38-4] (179) for the coconut rhinoceros beetle, *Oryctes rhinoceros*.

Other food lures which have had practical use in trapping insect pests include isoamyl salicylate [87-20-7] for moths of the tomato and tobacco hornworms, *Manduca* spp.; heptyl butyrate [5870-93-0], for stinging yellowjackets, *Vespula* spp.; and 1-octene-3-ol [3391-86-4], for the bloodsucking tsetse flies, *Glossina* spp.

Insecticide Residues in Foods

Approximately 70% of all insecticide use is in agriculture, and applications are generally made directly to raw agricultural commodities to protect plants and animals from insect attacks. With the exception of microbial insecticides, nearly all of these uses of insecticides result in residues of the various chemicals and their degradation products, which may be present in detectable amounts, ppb to ppm in food, despite weathering during crop maturation and attenuation during food processing. The World Health Organization (WHO) defines a pesticide residue as "any substance...resulting from the use of a pesticide that includes any specified derivatives such as degradation products, metabolites, and reaction products which are of toxicological significance." With increasing use of pesticides and the proliferation of plant protection chemicals, the nature and magnitude of such persisting residues has assumed great significance in public health and in the economics of international commerce in food products.

Residue Persistence Curves. The magnitude of persistence of pesticide residues on raw agricultural commodities is a function of the properties of the pesticide, the dosage applied, the structure of the surface of the treated plant, and the environmental conditions (weathering). The residue content is determined at intervals after application in mg/kg of weight of produce. These data generally follow first-order kinetics and can be analyzed by linear regression analysis. Such linear residue persistence curves are used to estimate the pesticide residue half-life which is independent of the rate of application. These residue persistence curves define the required interval after application to attenuate to a toxicologically safe level, ie, the maximum residue limit. This defines the safe interval before harvest as expressed on the pesticide label directions.

Toxicology of Insecticide Residues. Risk assessment from the chronic ingestion of insecticide residues is made from the results of lifetime feeding studies at several dosage ranges with mice, rats, and dogs. The observed no adverse effect level (NOAEL) in mg of pesticide ingested per kg of body weight is evaluated by considerations of the test animals' general health, food intake, weight gain, gross histopathology, blood chemistry, and enzyme activity. The acceptable daily intake (ADI) for humans exposed to pesticide residues in the diet is determined from such laboratory animal investigations by incorporating a safety factor to accommodate for inter- and intraspecific variations. This safety factor is set at 10-fold where valid human exposure data are available, at 100-fold where there are valid laboratory data but no human data, and at 1000-fold where no adequate chronic exposure data are available. Thus the ADI in mg per kg per day is an estimate of the daily pesticide dietary intake that appears to be without risk over the entire human lifetime. ADI values are established and periodically reviewed by joint committees of the Food and Agricultural Organization (FAO) and WHO of the United Nations.

The basic guidelines for the limitations of pesticide residues in raw agricultural commodities are the maximum residue limits (MRL) or residue tolerances established by national and international agencies. The MRL values represent the smallest amounts of pesticides that are practical and toxicologically acceptable when the pesticides are used according to good agricultural practices in the minimum quantities necessary to achieve adequate pest control. This definition allows considerable latitude and various countries have established differing tolerances for similar pesticide uses. International standardization, however, is highly desirable for international trade in foodstuffs, and a joint FAO–WHO Codex Committee has established and periodically reviews international MRL values for more than 100 pesticides used on more than 1600 specific food crops. In the United States, the establishment of pesticide tolerances is the responsibility of the Environmental Protection Agency under the Federal Insecticide, Fungicide, and Rodenticide Act of 1972. This provides "that a pesticide may be registered if it is effective, properly labeled, and when properly used will not cause unreasonable adverse effects on the environment, taking into account the economic, social, and environmental costs and benefits of the use of any pesticide."

BIBLIOGRAPHY

"Insecticides" in *ECT* 1st ed., Vol. 7, pp. 881–908, by D. E. H. Frear, The Pennsylvania State College, and M. A. Coler and M. H. Peskin, Markite Co.; in *ECT* 2nd ed., Vol. 11, pp. 677–738; "Insect Control Technology" in *ECT* 3rd ed., Vol. 13, pp. 413–485, by R. L. Metcalf, University of Illinois.

1. R. L. Metcalf and R. A. Metcalf, *Destructive and Useful Insects*, 5th ed., McGraw-Hill Book Co., Inc., New York, 1993.
2. H. H. Cramer, *Pflanzenschutz-Nachr.* **20** (1967).
3. R. L. Metcalf and J. J. McKelvey, Jr., eds., *The Future for Insecticides*, John Wiley & Sons, Inc., New York, 1976.
4. R. L. Metcalf, *J. Agr. Food Chem.* **21**, 511 (1973).
5. D. Pimentel and H. Lehman, eds., *The Pesticide Question*, Chapman & Hall, New York, 1993.
6. National Academy of Sciences, *Insect Pest Manage. Control*, **1695** (1969).
7. R. L. Metcalf, *Annu Rev. Entomol.* **25**, 219 (1980).
8. C. B. Huffaker, ed., *New Technology of Pest Control*, John Wiley & Sons, Inc., New York, 1980.
9. R. L. Metcalf and W. H. Luckmann, eds., *Introduction to Insect Pest Management*, 3rd ed., John Wiley & Sons, Inc., New York, 1994.
10. L. Pedigo, *Entomology and Pest Management*, Macmillan Publishing Co., New York, 1989.
11. E. Y. Spencer, *Guide to the Chemicals Used in Crop Protection*, 7th ed., Research Branch Agriculture, Ontario, Canada, 1980.
12. *Merck Index*, 11th ed., Merck & Co., Inc., Rahway, N.J., 1989.
13. *Agrochemicals Handbook*, Royal Society Chemistry, Cambridge, U.K., 1991.
14. R. L. Metcalf, *Organic Insecticides*, Interscience Publishers, New York, 1955.
15. R. D. O'Brien, *Insecticides: Action and Metabolism*, Academic Press, Inc., New York, 1967.
16. F. Matsumura, *Toxicology of Insecticides*, Plenum Press, New York, 1975.
17. K. H. Büchel, ed., *Chemistry of Pesticides*, Wiley-Interscience, New York, 1983.
18. C. F. Wilkinson, ed., *Insecticide Biochemistry and Physiology*, Plenum Press, New York, 1976.
19. D. E. H. Frear, *Chemistry of the Pesticides*, 3rd ed., D. Van Nostrand, Inc., Princeton, N.J., 1955.
20. M. Jacobsen and D. G. Crosby, eds., *Naturally Occurring Insecticides*, Marcel Dekker, New York, 1973.
21. K. Naumann, *Synthetic Pyrethroid Insecticides: Structures and Properties*, Springer-Verlag, Berlin, 1990.
22. P. Müller, ed., *DDT Insektizid Dichlorodiphenyltrichloroäthan und Seine Bedeutung*, Birkhauser, Basel, Switzerland, 1958.
23. G. T. Brooks, *Chlorinated Insecticides*, Vols. 1 and 2, CRC Press, Cleveland, Ohio, 1974.
24. J. R. Coats, ed., *Insecticide Mode of Action*, Academic Press, Inc., New York, 1982.
25. R. L. Metcalf, *Bull. Entomol. Soc. Am.* **5**, 3 (1959).
26. R. D. O'Brien, *Toxic Phosphorous Esters*, Academic Press, Inc., New York, 1960.
27. G. Schrader, *Die Entwicklung Neuer Insektizider Phosphorsäure-Ester*, 3rd. ed., Vedag Chemie, Weinheim, Germany, 1963.

28. M. Eto, *Organophosphorus Pesticides: Organic and Biological Chemistry*, CRC Press, Cleveland, Ohio, 1974.

29. C. Fest and K-J Schmidt, *The Chemistry of Organophosphorus Pesticides*, 2nd ed., Springer-Verlag, New York, 1980.

30. World Health Organization (WHO) Environmental Health Criteria, *Organophosphorus Insecticides*, Geneva, Switzerland, 1986.

31. M. K. Johnson, *CRC Crit. Rev. Toxicol.* **3**, 289 (1975).

32. J. M. Cramer and E. J. Hixon, eds., *Delayed Neurotoxicity*, Intox Press, Little Rock, Ark., 1984.

33. T. R. Fukuto, *Bull. WHO* **44**, 31 (1971).

34. T. Namba, *Bull. WHO* **44**, 289 (1971).

35. M. J. Kolbezen, R. L. Metcalf, and T. R. Fukuto, *J. Agr. Food Chem.* **13**, 220 (1954).

36. H. Gysin, *Chimia* **8**, 205 (1954).

37. R. L. Metcalf and T. R. Fukuto, *J. Agr. Food Chem.* **13**, 220 (1965).

38. R. J. Kuhr and H. W. Dorrough, *Carbamate Insecticides, Chemistry, Biochemistry, and Toxicology*, CRC Press, Cleveland, Ohio, 1976.

39. WHO Environmental Health Criteria, *Carbamate Pesticides*, Geneva, Switzerland, 1986.

40. L. I. Gilbert, ed., *The Juvenile Hormones*, Plenum Press, New York, 1976.

41. W. S. Bowers, *Bull. WHO* **44**, 381 (1971).

42. G. B. Stahl, *Annu. Rev. Entomol.* **20**, 417 (1975).

43. R. Wegler, ed., *Chemie der Pflanzenschutz- und Schadlingsbekämpfungsmittel*, Vol. 6, Springer-Verlag, Berlin, 1981.

44. A. Verdoop and C. D. Ferrell, in J. R. Plimmer, ed., *Pesticide Chemistry for the 20th Century*, American Chemical Society, Washington, D.C., 1977.

45. H. Schmutterer, *Annu. Rev. Entomol.* **35**, 27 (1990).

46. R. A. Dybas, in J. Miyamoto, ed., *Pesticide Chemistry: Human Welfare and the Environment*, Pergamon Press, Inc., New York, pp. 88–90, 1983.

47. D. L. Lindgren and L. C. Vincent, *Adv. Pest Control Res.* **5**, 185 (1962).

48. E. A. Steinhaus, *Insect Pathology: an Advanced Treatise*, Vols. 1 and 2, Academic Press, Inc., New York, 1973.

49. E. Kurstak, ed., *Microbial and Viral Pesticides*, Marcel Dekker, New York, 1982.

50. C. M. Ignoffo, ed., *Handbook of Naturally Occurring Pesticides: Microbial Insecticides*, CRC Press, Boca Raton, Fla., 1986.

51. J. V. Maddox, in S. B. Vinson and R. L. Metcalf, eds., *Entomology Serving Society: Emerging Technologies and Challenges*, Entomology Society of America, Lanham, Md., 1991, p. 356.

52. A. W. A. Brown and R. Pal, *Insecticide Resistance in Arthropods*, WHO, Geneva, Switzerland, 1971.

53. G. P. Georghiou and T. Saito, eds., *Pest Resistance to Pesticides*, Plenum Press, New York, 1984.

54. National Academy of Sciences, *Pesticide Resistance: Strategies and Tactics for Management*, Washington, D.C., 1986.

55. R. L. Metcalf, *Pest. Sci.* **26**, 333 (1989).

56. R. T. Roush and B. E. Tabashnik, eds., *Pesticide Resistance in Arthropods*, Chapman & Hall, New York, 1990.

57. G. P. Georghiou and A. Lagunes-Tejeda, *The Occurrence of Resistance to Pesticides in Arthropods*, FAO, Rome, 1991.

58. E. F. Knipling, *The Basic Principles of Insect Population Suppression and Control*, U.S. Dept. Agricultural Handbook No. 512, Washington, D.C., 1976.

59. M. A. Hoy and J. J. McKelvey, Jr., *Genetics in Relation to Insect Management*, Rockefeller Foundation, New York, 1979.

60. R. Pal and M. J. Whitten, eds., *The Use of Genetics in Insect Control*, Elsevier, Amsterdam, the Netherlands, 1974.

61. A. B. Borkovic, *Insect Chemosterilants*, Wiley-Interscience, New York, 1966; G. C. LaBreque and C. N. Smith, *Principles of Chemosterilization*, Elsevier, Amsterdam, the Netherlands, 1968.

62. V. Dethier, *Annu. Rev. Entomol.* **1**, 181 (1956).

63. G. F. Shambaugh, R. F. Brown, and J. J. Pratt, Jr., *Adv. Pest Control Res.* **1**, 277 (1957).

64. S. A. Hall, N. Green, and M. Beroza, *J. Agr. Food Chem.* **5**, 66 (1957).

65. C. N. Smith, *Repellents for Anopheline Mosquitoes*, *Misc. Pub. Entomol. Soc. Am.* **7**, 99 (1970).

66. D. T. Waterhouse, *Adv. Pest Control Res.* **2**, 207 (1958).

67. H. H. Shorey and J. J. McKelvey, Jr., eds., *Chemical Control of Insect Behavior*, Wiley-Interscience, New York, 1977.

68. E. R. Mitchell, ed., *Management of Insect Pests with Semiochemicals*, Plenum Press, New York, 1981.

69. D. A. Nordlund, R. L. Jones, and W. J. Lewis, eds., *Semiochemicals: Their Role in Pest Control*, Wiley-Interscience, New York, 1981.

70. R. L. Ridgway, R. M. Silverstein, and M. N. Inscoe, *Behavior-Modifying Chemicals for Insect Management*, Marcel Dekker, New York, 1990.

71. R. L. Metcalf and E. R. Metcalf, *Plant Kairomones in Insect Ecology and Control*, Chapman & Hall, New York, 1992.

72. G. Vettorazzi, *CRC Crit. Rev. Toxicol.* **4**, 125 (1975).

73. National Academy of Sciences, *Regulating Pesticides*, Washington, D.C., 1980.

74. G. Vettorazzi, *International Regulatory Aspects of Pesticide Chemicals*, Vols. 1 and 2, CRC Press, Cleveland, Ohio, 1979.

75. C. M. Menzie, *Metabolism of Pesticides*, U.S. Dept. Interior Special Scientific Report Wildlife, Washington, D.C., 1966, 1974, and 1978, nos. 96, 184, 212.

76. H. Aizawa, *Metabolic Maps of Pesticides*, Academic Press, Inc., New York, 1982.

77. N. A. Smart, *Residue Rev.* **98**, 99 (1985).

Robert L. Metcalf
University of Illinois, Urbana-Champaign

INSTRUMENTATION AND CONTROL

. See PROCESS CONTROL.

INSULATION, ACOUSTIC

Acoustic insulation may be defined as a material or construction that reduces the passage or transmission of sound into or out of a medium such as air, water, or a solid structure. For the purpose of this article, sound is a vibratory disturbance in the air consisting of alternating compressions and rarefactions transmitted through the air in waves. The term acoustic insulation covers a broad range of materials and mechanisms for the control of sound: sound-absorbing materials that reduce the reflections of impinging sounds; sound-blocking materials that reduce sound transmission from one location to another; vibration isolating materials and devices that reduce transmission of vibrations from a vibrating source to potential sound-radiating structures; and vibration damping materials that reduce vibrations and sound radiation in and from materials and structures.

Sound Absorption

When a sound wave strikes a material a fraction of its energy is reflected and a fraction is dissipated, or absorbed, by the material. The fraction of sound energy absorbed by a material is designated by its sound-absorption coefficient (α). The sound-absorption coefficient of a given material is between zero and one; if it is zero all the impinging energy is reflected and none absorbed; if it is one all the energy is absorbed and none reflected.

Units. The unit of sound absorption is the metric sabin, which is equivalent to one square meter of "perfect" absorption, eg, one square meter of a material with $\alpha = 1.0$. The English unit of sound absorption is the sabin, which is equivalent to one square foot of perfect absorption. In order to avoid confusion, the designation metric should always be used when referring to metric sabins. The number of metric sabins of absorption provided by an area of material is calculated by multiplying its area by its sound-absorption coefficient. For example, 10 m² of material having a sound-absorption coefficient of 0.75 provides 7.5 metric sabins of absorption.

The sound absorption of materials is frequency dependent; most materials absorb more or less sound at some frequencies than at others. Sound absorption is usually measured in laboratories in 18 one-third octave frequency bands with center frequencies ranging from 100 to 5000 Hz, but it is common practice to publish only the data for the six octave band center frequencies from 125 to 4000 Hz. Suppliers of acoustical products frequently report the noise reduction coefficient (NRC) for their materials. The NRC is the arithmetic mean of the absorption coefficients in the 250, 500, 1000, and 2000 Hz bands, rounded to the nearest multiple of 0.05.

All materials, even those considered to be sound-reflecting, absorb some small fraction of the sound energy impinging on them. Table 1 provides sound-absorption coefficients for some common building materials.

Table 1. Sound-Absorption Coefficients (α) for Some Common Building Materials

Material	Octave band center frequency, Hz						NRC
	125	250	500	1000	2000	4000	
brick	0.03	0.03	0.03	0.04	0.05	0.07	0.05
concrete block							
coarse	0.36	0.44	0.31	0.29	0.39	0.25	0.35
painted	0.10	0.05	0.06	0.07	0.09	0.08	0.05
concrete, smooth	0.01	0.01	0.02	0.02	0.02	0.03	0.05
gypsum board on wood studs	0.29	0.10	0.05	0.04	0.07	0.09	0.05
heavy plate glass	0.18	0.06	0.04	0.03	0.02	0.02	0.05
wood floor	0.15	0.11	0.10	0.07	0.06	0.07	0.10

Test Methods. Two basic types of test methods are commonly used to measure sound-absorption in test laboratories: the reverberation room method and the impedance tube method.

Reverberation Room Test Method. The more widely used test method is the reverberation room method, defined in the United States by the American Society for Testing and Materials (ASTM) C423-90a (1). The basis for this test is the relationship that the rate of decay of an instantaneously stopped sound in a room is proportional to the amount of sound absorption in the room. The material is tested in a reverberation room where all other surfaces are hard and sound-reflecting. The rate of decay in each frequency band is measured with and without the sample, the number of metric sabins contributed by the sample is calculated, and the sound-absorption coefficients are determined based on the size of the sample and the amount of absorption provided. Because of edge effects the calculated absorption coefficients for a small sample are larger than for a large sample of the same material. To minimize this problem the standard recommends a sample size of 6.69 m² (72 ft²) and requires that it not be less than 4.46 m² (48 ft²). Even for the recommended sample size the measured sound absorption is influenced by edge effects, and for very efficient sound-absorbing materials the result can be calculated absorption coefficients greater than 1.0. These high coefficients, which are sometimes reported by manufacturers of acoustical products, are artifacts of the test procedure. Sound-absorption coefficients greater than one should never be used for acoustical analysis purposes.

The sound-absorbing properties of acoustical materials also are influenced by the manner in which the materials are mounted. Standard mounting methods for use in laboratory testing are specified in ASTM E795-92 (2). Unless noted otherwise, published data for acoustic ceiling materials are for Mounting Type E-400, for which the material being tested is suspended 400 mm below a hard surface.

For reverberation room tests of some irregularly shaped items, such as items of furniture, the number of sabins of absorption per item is commonly reported, rather than the absorption coefficient. It is important that the number and arrangement of the items also be reported because both of these factors can affect the results of the test.

Because the reverberation room test method approximates many real-world conditions, it is used to derive sound-absorption coefficients for evaluating the effect of most actual applications of sound-absorbing treatments. Sound-absorption coefficients published in acoustical textbooks and by manufacturers of acoustical materials are almost exclusively from reverberation room tests, and this may be assumed unless specified otherwise.

Impedance Tube Test Methods. There are two impedance tube test methods: ASTM C384-90a (3) and ASTM E1050-90 (4). Test method C384-90a makes use of a tube with a test specimen at one end, a loudspeaker at the other, and a probe microphone that can be moved inside the tube. Sound emitted from the loudspeaker propagates down the tube and is reflected back by the specimen. A standing wave pattern develops inside the tube,

and the probe microphone determines the nature of the pattern. The normal incidence sound-absorption coefficient (α_n) is then calculated by means of a formula based on the ratio of the maximum to minimum pressure of the standing wave.

ASTM E1050-90 also makes use of a tube with a test specimen at one end and a loudspeaker at the other end, but instead of a single movable microphone there are two microphones at fixed locations in the tube. The signals from these microphones are processed by a digital frequency analysis system which calculates the standing wave pattern and the normal incidence sound-absorption coefficients.

One advantage of the impedance tube test methods is the small (usually <10 cm (4 in.) dia) size of the test samples. For these tests sound impinges on the test sample only at normal incidence to the surface, and the sound-absorption coefficients derived in this manner are valid only at this angle. Because of this limitation the tube methods are used primarily for research and for applications where sound is incident only normal to the surface of a material.

Materials.

Fibrous and Foamed Materials. Most sound-absorbing materials are fibrous or porous and are easily penetrated by sound waves. Air particles excited by sound energy move rapidly to and fro within the material and rub against the fibers or porous material. The frictional forces developed dissipate some of the sound energy by converting it into heat.

The fibrous materials most often used for sound-absorbing purposes are composed of either glass fibers or mineral fibers. Fibrous glass, commonly known as fiber glass, is manufactured by forcing molten glass through a series of nozzles to form liquid fibers that are then split into smaller fibers with diameters of 1 to 10 μm (see GLASS). The fibers are formed into unfinished blankets, batts, and boards of various densities, using a variety of binders to hold the fibers together. Glass fibers tend to shred or settle on contact or in the presence of vibration or high velocity air flow, so unfinished fiber glass products are often used for acoustical purposes behind protective sound-transparent facings. Fiber glass used for acoustical purposes should not be confused with fiber glass-reinforced plastic materials, also commonly referred to as fiber glass, which are used in boat construction and other products.

Rock wool, frequently referred to as mineral fiber, is made from nonvirgin siliceous materials and is formed in a similar manner to that of fiber glass. Refractory fibers (qv), also formed in a similar manner, are available for high temperature applications.

Foamed plastic acoustical materials are manufactured by two different processes. Both processes involve combining reactants that simultaneously produce a polymer, typically polyurethane, and generate a gas. Bubbles of gas expand the reacting mass and eventually form contiguous polyhedrons. If the contact planes between the polyhedrons rupture and establish openings between the cells, allowing air to penetrate, the material will have useful sound-absorbing properties. On the other hand, if the cells remain closed so air cannot penetrate, the foamed material will be ineffective as a sound-absorbing treatment. In one process a reacting mass is batch-formulated by allowing it to form a large bun about 1 meter thick that is then cut into sheets. Pressure-sensitive adhesives, mass-loaded backings or septums, or thin impervious plastic film facings are sometimes applied to the sheets. In another process the foam product is continuously cast and formed into a final thickness, and various substrates are applied as the foam is formed (see FOAMED PLASTICS).

Other fibrous and porous materials used for sound-absorbing treatments include wood, cellulose, and metal fibers; foamed gypsum or Portland cement combined with other materials; and sintered metals. Wood fibers can be combined with binders and flame-retardant chemicals. Metal fibers and sintered metals can be manufactured with finely controlled physical properties. They usually are made for applications involving severe chemical or physical environments, although some sintered metal materials have found their way into architectural applications. Prior to concerns regarding its carcinogenic properties, asbestos fiber had been used extensively in spray-on acoustical treatments.

Resonant Sound Absorbers. Two other types of sound-absorbing treatments, resonant panel absorbers and resonant cavity absorbers (Helmholtz resonators), are used in special applications, usually to absorb low frequency sounds in a narrow range of frequencies. Resonant panel absorbers consist of thin plywood or other membrane-like materials installed over a sealed airspace. These absorbers are tuned to specific frequencies, which are a function of the mass of the membrane and the depth of the airspace behind it. Resonant cavity absorbers consist of a volume of air with a restricted aperture to the sound field. They are tuned to specific frequencies, which are a function of the volume of the cavity and the size and geometry of the aperture.

Uses. Sound-absorbing materials are frequently used to reduce reverberation, or the persistence of sound in a space after generation of the sound ceases; to reduce focused reflections from concave surfaces; to prevent echoes, or delayed sound reflections from distant surfaces; and to prevent the buildup of sound by multiple reflections within rooms and other enclosures. Sound-absorbing materials also are used to reduce the transmission of noise from one location to another by multiple reflections from sound-reflecting surfaces.

Reverberation Control. Reverberation time (T_{60}) is defined as the length of time in seconds for the sound of an instantaneously stopped source in a room to decay by 60 decibels (dB). Reverberation time is one important factor in determining the acoustical character of a space and its suitability for specific activities. For lectures and other speech activities a relatively short reverberation time is desirable so that syllables do not persist and overlap one another, causing difficulty with intelligibility; conversely, for music activities, a relatively long reverberation time is desirable to allow blending of the sound and a sense of being surrounded by the music. Without reverberation music usually sounds dull and lifeless.

The reverberation time in a room is directly proportional to the volume and inversely proportional to the amount of sound absorption in the room. For most practical purposes the reverberation time is determined by the Sabine equation:

$$T_{60} = \frac{0.161 V}{A}$$

where V is the volume of the room in m^3 and A is the total absorption in the room in metric sabins (5). Thus the reverberation time in an existing room can be decreased by adding sound absorption, or increased by removing sound absorption; if the amount of sound absorption is doubled, the reverberation time is cut in half, and vice versa. More sophisticated equations are sometimes used to take into account air absorption and other acoustical effects not accounted for in the Sabine equation, but for many applications the Sabine equation provides a satisfactory degree of accuracy.

Noise Reduction in Rooms. Sound from a source in an enclosed room can be divided into two parts: the direct field, dominated by sound radiated directly from a source to a receiver without reflections; and the reverberant field, dominated by sound that has been reflected many times by surfaces in the room before it reaches a receiver (5). This relationship is defined by the following equation:

$$L_p(r) = L_w + 10 \log_{10} \left(\frac{Q_\phi}{4 \pi r^2} + \frac{4}{R} \right)$$

where $L_p(r)$ is the sound pressure level (dB at 20 μPa) at a distance r from the sound source and away from the immediate vicinity of any reflecting surfaces; L_s is the sound power level emitted by the source into the space (dB at 10⁻¹² W); Q_ϕ is the directivity factor of the source in the direction ϕ , R is the room constant, m^2 ; $Q_\phi / 4 \pi r^2$ represents the direct field; and $4 / R$, represents the reverberant field. The room constant, R , is difficult to determine, and for practical purposes the total absorption, A , may be substituted for R . Sound in the direct field is a function of distance from the source and drops off at approximately 6 dB per doubling of distance. Sound in the reverberant field is primarily a function of the amount of absorption in the room; each doubling of the amount of absorption reduces the reverberant sound pressure level by 3 dB. For most existing untreated rooms the practical upper limit of reduction that can be achieved for remedial purposes is about 10 dB. Figure 1 is a plot of the relative sound pressure level vs distance for spaces with total sound absorption ranging from 50 to 10,000 metric sabins. The sloped portions of the curves represent the sound in the near field of the source, while the flat portions represent the sound in the reverberant field.

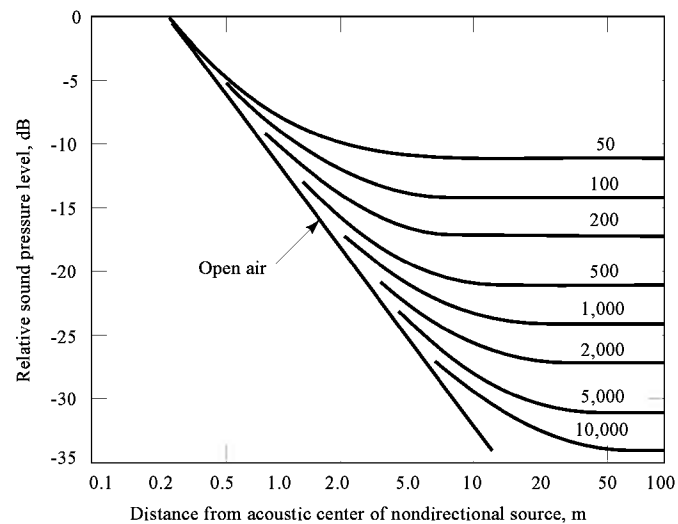


Fig. 1. Approximate relationship between sound pressure level and distance as a function of room absorption. Numbers on the curves are metric sabins.

Noise Transmission Reduction in HVAC Systems. One common use of sound-absorbing treatment is to reduce noise transmission in heating, ventilating, and air-conditioning (HVAC) systems (6). The treatments are used to reduce the transmission of fan noise and air turbulence noise through ducts into occupied spaces. Noise transmission reduction in duct systems is described in terms of insertion loss, the difference in sound power level or sound pressure level measured at a given location before and after installation of the treatment; or sound attenuation, the reduction in sound power between two locations affected by a sound source. The units are decibels.

Sound Blocking. Most sound-absorbing materials by themselves are not effective as sound barriers for blocking sound transmission; air can penetrate these materials and so can sound. As a result, wrapping or enclosing noisy equipment in fiber glass or porous foam materials does little to reduce the noise radiated to the surrounding areas. Although sound-absorbing materials by themselves are ineffective as sound barriers, they can improve the sound-isolating performance of impervious materials when combined with them.

Products. There is a large number of commercially available sound-absorbing products for use on ceilings, walls, and for other special applications. Sound absorption coefficients and NRC values for some sound-absorbing products and treatments are indicated in Table 2.

Table 2. Sound-Absorption Coefficients (α) for Some Sound-Absorbing Treatments

Treatment	Octave band center frequency, Hz						NRC
	125	250	500	1000	2000	4000	
1.9-cm thick mineral fiber acoustic tiles ^a							
low absorption	0.40	0.30	0.54	0.78	0.67	0.48	0.55
high absorption	0.67	0.62	0.66	0.88	0.99	0.99	0.80
2.5-cm nubby fiber glass ceiling panels	0.70	0.95	0.75	0.99	0.99	0.99	0.90
fabric-wrapped fiber glass panels							
2.5-cm thick	0.07	0.37	0.73	0.97	0.99	0.99	0.75
5.1-cm thick	0.23	0.81	0.99	0.99	0.99	0.99	0.95
heavy velour draperies ^b	0.15	0.35	0.55	0.70	0.70	0.70	0.60
thin carpet on concrete	0.02	0.05	0.10	0.15	0.25	0.50	0.15
heavy carpet on pad on concrete	0.05	0.10	0.30	0.50	0.70	0.80	0.40
7.6-cm acoustical steel deck ^c	0.73	0.99	0.99	0.89	0.52	0.31	0.85
2.5-cm sprayed cellulose fiber	0.08	0.29	0.75	0.98	0.93	0.76	0.75

^a Mounting E-400 (Ref. 2).

^b 50° o fullness spaced 15 cm from hard surface.

^c Ribbed deck with ribs filled with fiber glass and sides perforated (Fig. 2).

Unfinished Products. Unfinished fiber glass products are available in the form of boards, blankets, and batts in various thicknesses and densities. These products are used by fabricators who apply finishes to make products suitable for ceilings, walls, open-plan office screens, etc. They also are used for sound absorption behind decorative and protective facings such as perforated or expanded metal and wood grilles. Thicker materials have better low frequency performance than thinner materials. Low frequency performance can be improved by spacing the material away from a sound-reflecting surface rather than applying the material directly to the surface.

Ceiling Tiles and Panels. Acoustical ceiling tiles and lay-in panels are the most commonly used acoustical products for noise and reverberation control in architectural applications. The majority of ceiling tiles and panels are manufactured using mineral fibers and a gypsum binder, although fiber glass panels also are common. The units are factory painted using nonbridging paints. Repainting, which can bridge the openings that allow sound to penetrate and be absorbed, should be done carefully and only with nonbridging paints. Typical tiles are 30.5 cm (12 in.) by 30.5 cm and range in thickness from 1.3 cm (S in.) to 1.9 cm (1 3 in.). They are installed most frequently as suspended ceilings below the structural members and ventilating ducts in commercial buildings, although they also can be adhesive-applied to gypsum board or plaster ceilings. Most acoustical ceiling panels are 61 cm (24 in.) by 61 cm or 61 cm (24 in.) by 122 cm (48 in.), although other sizes are sometimes available. Thicknesses range from 1.3 cm (S in.) to 3.8 cm (1-S in.). The panels usually are laid into a grid of horizontal members having an inverted T-shaped section, allowing the panels to be lifted for access to the plenum space above the ceiling. Ceiling tiles and panels are generally soft and friable and are not suitable for use on surfaces that are subject to abuse.

Metal Pan Assemblies. These units consist of tiles and panels formed from perforated aluminum or steel with pads of fiber glass or mineral wool inserted into the pans to provide the sound absorption. They are used primarily for ceilings in a similar manner to acoustical tiles and panels. The pads are sometimes sealed in plastic film to prevent absorption of moisture, dirt, and odors. The perforated metal is relatively sound transparent and functions as the finished ceiling and the support for the sound-absorbing material. The perforated metal by itself has no acoustical value.

Spray-On Treatments. Several types of sound-absorbing treatments are available for spray-on application to a backup surface of concrete, gypsum board, plaster, or other hard and reflective material. Gypsum, Portland cement, and cellulose-based materials are used in these applications, which employ a spray process that generates a rigid foam-like structure with interconnecting pores. The procedure, equipment, and some of the products are similar to those employed in spray-on fireproofing of building structures. They are commonly sprayed to thicknesses ranging from about 1.3 cm (S in.) to 5 cm (2 in.), with the thicker treatments typically providing greater sound absorption. Some of these treatments are known as acoustical plasters and their surfaces are sometimes modified by troweling or screeding. A problem with spray-on treatments is that the acoustical performance depends on careful control of the application procedure and thickness, so the desired acoustical performance may not always be achieved. Sprayed asbestos was a popular and

effective sound-absorbing treatment prior to concerns regarding its carcinogenic properties.

Acoustical Roof Decks. Acoustical roof decks are frequently used in gymnasiums, one-story commercial buildings, and similar utilitarian rooms and buildings. They comprise part of the roof structure, taking the place of nonacoustical roof decking materials. Acoustical metal roof decks consist of structural panels that are hollow or ribbed and are filled with fiber glass or mineral wool sound-absorbing material; the bottom faces of the deck or the sides of the ribs are perforated metal. One common type of acoustical metal deck is illustrated in Figure 2. Another type of acoustical roof deck is made with shredded wood fibers with a gypsum or Portland cement binder. Conventional insulation and roofing are installed on top of the acoustical roof decks.

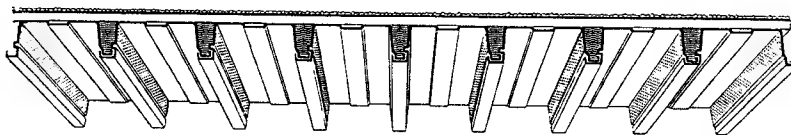


Fig. 2. Typical ribbed acoustical metal deck.

Courtesy of Inryco.

Acoustical Wall Treatments. Sound-absorbing wall treatments are sometimes required to prevent echoes (long-delayed reflections) or flutter echoes (repeated reflections between hard parallel surfaces). Prefabricated fiber glass boards 2.5 cm (1 in.) to 5.1 cm (2 in.) thick wrapped in fabric or thin perforated vinyl are widely used. Mineral fiber boards with integral facings of fabric-like material also are available. Open-cell foam products represent another type of acoustical treatment used on walls. The foam products are available in several thicknesses and with sculptured surfaces, some of which have a sawtooth shape and others a textured appearance similar to egg cartons. They can be applied to wall surfaces by means of a self-adhesive backing and are popular in sound studios because of their appearance as well as their properties. Porous concrete blocks with carefully designed slots and cavities, which act as resonant cavity absorbers, also are used for sound-absorbing walls. The resonant cavities provide a significant amount of low frequency absorption, and the porous concrete provides some middle and high frequency absorption.

Custom decorative sound-absorbing treatments for wall surfaces are frequently used in auditoriums and theaters, especially for control of echoes from rear walls. Typical treatments consist of prefabricated or custom-built wood grilles over fiber glass or mineral wool blankets or batts.

Carpeting. Carpeting may be used for noise and reverberation control, but because of its limited thickness it is only effective at relatively high frequencies. Thick pile carpeting is more effective than thin carpeting, and installation over foam pads further improves the sound-absorbing properties. When installed over a foam pad, carpeting should not have an impermeable backing, eg, latex, but should have an open back to allow sound to penetrate into the pad. In addition to providing sound absorption at higher frequencies, carpeting provides a resilient surface that reduces the noise of heel clicks, footfalls, and other sounds originating on the floor.

Draperies. Draperies of light weight or open-weave fabrics are ineffective for sound-absorbing purposes. Heavy draperies, such as flannel and velour, can provide useful sound absorption if properly installed. For best results they should be hung with 100% fullness, ie, 2 m² for every m² of wall or window surface covered. The sound-absorbing properties also are affected by the amount of space between the draperies and the surface behind them.

Unit Absorbers. Sound-absorbing baffles are the most common type of unit absorber. Typical baffles are 61 cm (24 in.) by 122 cm (48 in.) and consist of a 5.1 cm (2 in.) fiber glass core wrapped in thin plastic film, fabric, or perforated vinyl. The baffles usually are suspended from the ceiling by one of the long sides in rows or in the form of a grid. Baffles wrapped in thin plastic film are frequently used to provide sound absorption in factories and other industrial applications where suspended ceilings are impractical because of the large amount of suspended ductwork, conduits, and other mechanical and electrical equipment requiring periodic access. Fabric-wrapped baffles are used in open-plan offices, cafeterias, and other nonindustrial applications where appearance is an important factor. Cylindrical unit absorbers also are available with fabric or plastic film finishes, and units having a triangular or diamond-shaped cross section are sometimes fabricated from acoustical ceiling panels. Acoustical performance of unit absorbers is usually described in terms of sabins per unit rather than absorption coefficients.

Acoustical Duct Lining. Acoustical duct lining is used to reduce transmission of fan and air turbulence noise through heating and air-conditioning duct systems (6). Most duct lining products are made of low density fiber glass with special facings or treatment to resist erosion and moisture in the air stream. The most useful thickness is 2.5 cm (1 in.), although 1.3 cm (½ in.) and 5 cm (2 in.) thicknesses also are available. The performance of the 1.3 cm lining is restricted to higher frequencies and is not suitable for most applications. The 5 cm material provides better low frequency performance than the 2.5 cm material, but its use is generally limited because of the large duct sizes required. Two types of ducts with integral sound-absorbing treatment also are available. One type is formed of fiber glass boards with an outer facing of aluminum foil. Another type consists of double-wall round or oval ducts having fiber glass sound-absorbing material between a solid sheet metal outer wall and a perforated metal inner wall. These double-wall ducts can be used in high velocity systems where standard acoustical duct lining or fiber glass ducts would not withstand the high velocity air flow. The acoustical performance of duct systems is measured in decibels of sound attenuation per unit length.

Duct Silencers. Duct silencers, which make use of sound-absorbing materials and restrictive air passages to dissipate sound energy, are known as dissipative mufflers. Reactive mufflers, widely used in motor vehicles, do not use sound-absorbing materials. The most common type of duct silencer consists of a rectangular section of duct containing a number of perforated metal baffles filled with sound-absorbing material, usually fiber glass or mineral wool. Air and sound flow between the parallel baffles and sound is absorbed by the acoustical material. In applications where contamination of the air stream by the fibrous material is a concern, the fill material can be bagged in thin plastic film. In some critical "clean" installations, where even bagged fill is inappropriate, "packless" silencers using only resonant cavity absorption principles are available. Duct silencers are available in standard lengths of 0.9 m (3 ft), 1.5 m (5 ft), 2.1 m (7 ft), and 3.0 m (10 ft). Custom sizes also are available from some manufacturers. Acoustical performance varies with frequency and air flow, and is described in terms of insertion loss as a function of frequency for various forward and reverse air flow velocities. Silencers induce pressure drop in duct systems, and because this is an important parameter in HVAC system design, various silencer configurations are available that result in varying amounts of pressure drop and insertion loss. Higher performance silencers generally produce higher pressure drops. A typical duct silencer is illustrated in Figure 3 and its rated acoustical performance is represented in Table 3.

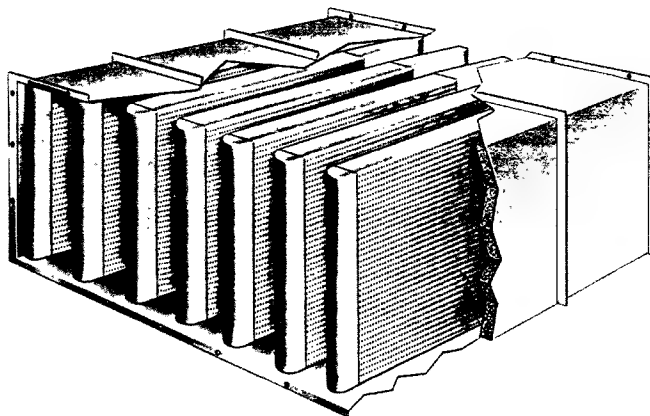


Fig. 3. Parallel baffle duct silencer; see Table 3.

Courtesy of Industrial Acoustics Co., Inc.

Table 3. Performance of a Typical Duct Silencer^a

Face velocity, ^b m min	Octave band center frequency, Hz ^c							
	63	125	250	500	1000	2000	4000	8000
1220	9	12	21	34	43	33	22	9
610	8	11	18	32	42	33	22	11
±610	6	10	18	30	42	34	23	14
±1220	4	9	17	29	38	34	23	14

^a Length 1.5 m.
^c Having dynamic insertion loss in dB.
^b Forward (±) and reverse () flow.

Acoustical Louvers. Acoustical louvers are used in building mechanical systems when exterior walls are penetrated for fresh air intake, exhaust, or relief air, in situations where the impact of HVAC noise is of concern in the surrounding environment. The louvers consist of a series of hollow sheet metal blades. The bottom faces of the louver blades are perforated and the blades are filled with fibrous sound-absorbing material. Typical acoustical louvers are 20 cm (8 in.) to 30 cm (12 in.) in depth. The amount of insertion loss they provide is limited.

Acoustical Lagging. Sound-absorbing materials are frequently used in combination with sound-blocking materials to reduce noise radiated from pipes, ducts, gearboxes, and other noise sources. This procedure, known as lagging (7), usually consists of 2.5 cm (1 in.) or more of acoustical insulation wrapped around the offending source, with a covering of sheet lead, mass-loaded vinyl, or some other heavy impervious material. A heavy outer covering is an essential part of the treatment, because sound easily passes through porous sound-absorbing materials. The sound-absorbing material provides some sound absorption as well as decoupling between the source and the outer jacket.

Sound Isolation

When a sound wave comes in contact with a solid structure, such as a wall between two spaces, some of the sound energy is transmitted from the vibrating air particles into the structure causing it to vibrate. The vibrating structure, in turn, transmits some of its vibrational energy into the air particles immediately adjacent on the opposite side, thereby radiating sound to the adjacent space. For an incomplete barrier, such as a fence or open-plan office screen, sound also diffracts over the top and around the ends of the barrier. The subject of this section is confined to complete barriers that provide complete physical separation of two adjacent spaces. Procedures for estimating the acoustical performance of partial barriers can be found in References 5 and 7.

Two useful measures of the performance of a sound-isolating construction are sound transmission loss (*TL*) and noise reduction (*NR*). Sound transmission loss is defined as follows, where *W_i* is the incident sound power (Watts) on the source side of the specimen, and *W_t* is the transmitted sound power on the receiving side (7).

$$TL = 10 \log_{10} (W_i / W_t)$$

Noise reduction (*NR*) is the difference in the average sound pressure level between the source room and the receiving room. When the receiving room is relatively reverberant and the measurements are made in the reverberant fields of the two rooms the relationship between *TL* and *NR* is as follows, where *S* is the surface area of the sound barrier between the two rooms and *A* is the amount of sound absorption in the receiving room (7).

$$TL = NR + 10 \log_{10} (S / A)$$

Units and Rating Procedures. The unit of sound pressure level is the decibel (dB), defined as follows where *L_p* is the sound pressure level, *p* is the measured sound pressure, and *p_{ref}* is the reference sound pressure of 20 μPa. *TL* and *NR* also are expressed in decibels.

$$L_p = 10 \log_{10} (p / p_{ref})$$

The sound-isolating performance of materials and structures vary with frequency. Sound-transmission loss is measured in one-third octave frequency bands with center frequencies ranging from 100 to 5000 Hz. In the past, the arithmetic mean of the one-third octave *TL* values (the average sound-transmission loss) was used to provide a single-number rating, but this number can be misleading and is now rarely used. A widely used single-number rating for laboratory measurements of sound-transmission loss is the sound-transmission class (STC). This rating is determined by comparing the measured sound-transmission loss curve with a reference curve that is moved up and down in level relative to the measured curve until certain criteria are met. The STC rating is then established by the level of the reference curve at 500 Hz. The procedure is described in ASTM E413-87 (8). Although the STC was developed for rating the performance of constructions for isolating speech, it is also frequently used for rating the overall performance of sound-isolating constructions. STC rating curves are illustrated in Figure 4.

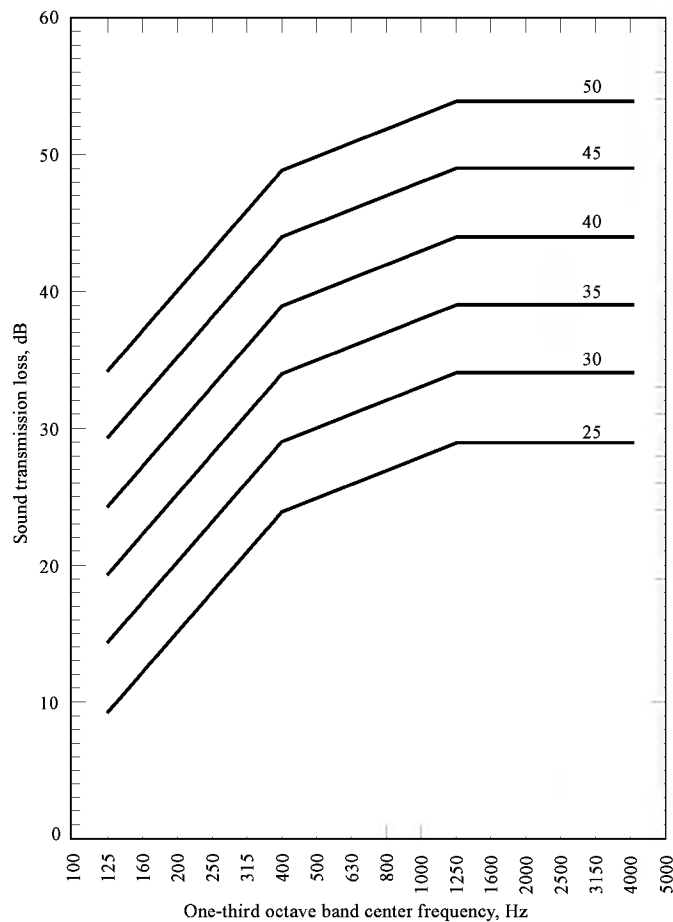


Fig. 4. Typical STC rating curves. Numbers on curves designate sound-transmission class (STC).

Two single-number ratings are used for field measurements of sound isolation: noise isolation class (NIC) and field sound-transmission class (FSTC). For NIC ratings the measured noise reduction between two rooms is rated using the procedure just described for STC ratings, with no corrections for receiving room absorption or other field irregularities. FSTC ratings, on the other hand, take into account the amount of absorption in the receiving room, and require complex procedures to ensure that there are no flanking paths around the sound-isolating element being measured and rated. Because of its simplicity, NIC is the more widely used of the two rating procedures for field measurements of sound isolation between rooms.

These procedures are used only for rating airborne sound isolation. A related procedure is used for rating the effectiveness of floor-ceiling constructions in reducing impact noise transmission, such as footsteps, from upper floors to rooms below. The noise produced by a standard "tapping machine" is measured in the room below, and is rated using a standard curve in a similar manner to the STC rating procedure. This procedure is described in ASTM E492-90 (9). The result is a single-number rating called the impact isolation class (IIC). A great deal of controversy exists over impact rating procedures and criteria. An older impact rating procedure, which is now obsolete, is the impact noise rating (INR). When INR was developed a rating of zero was considered to provide adequate impact isolation. An IIC rating of 51 is approximately equal to an INR of zero. Impact isolation criteria for multifamily dwellings were established in the 1960s by the U.S. Department of Housing and Urban Development (10). These remain the most comprehensive criteria of this type in the United States.

Test Methods.

Laboratory Methods. The laboratory test method for determining the sound-transmission loss performance of constructions is defined in ASTM E90-90 (11). The sample is installed in an opening between two highly reverberant rooms that are acoustically well isolated from each other. Rotating vanes are provided in the rooms to ensure diffuse sound fields. Sound is introduced into the source room, the average sound pressure level is measured in one-third octave bands in both rooms, and the sound-transmission loss is calculated as follows, where $\bar{L}_1$ and $\bar{L}_2$ are the average sound pressure levels in the source and receiving rooms; S is the area of the test sample, m²; and A_2 is the absorption in the receiving room, metric sabins.

$$TL = \bar{L}_1 - \bar{L}_2 + 10 \log S - 10 \log A_2$$

Field Methods. The purpose of noise reduction measurements in buildings is to determine the overall sound-isolating performance of the construction. Random noise is introduced into a source room. The space-averaged noise is then measured in the source room and the receiving room in one-third octave or octave frequency bands. The noise reduction is determined by subtracting the measured sound pressure levels in the receiving room from those in the source room. The noise isolation class (NIC) is determined using the STC rating procedure. Measurement of field sound-transmission loss is similar to noise reduction, except that it is used to determine the sound-isolating performance of a single element of the construction and can be compared directly to laboratory measurements. It may be necessary to construct barriers to shield other elements of the construction to ensure that they do not contribute to the measured sound levels in the receiving room. In addition, the amount of sound absorption must be determined in the receiving room in order to convert the measured noise reduction to transmission loss. The STC rating procedure is used to determine the field sound-transmission class (FSTC). The field test method is defined in ASTM E336-90 (12).

Materials. All common building materials provide some degree of sound isolation when used to separate adjacent spaces. The sound-isolating performance depends on a number of factors including mass, stiffness, size, and complexity of construction. In general, materials used for sound-isolating purposes must be impervious to air penetration; therefore porous materials like fiber glass and rock wool, which air can penetrate, are not effective for sound-isolating purposes unless combined with impervious materials. Heavy materials, such as concrete or lead, provide more sound isolation than lighter ones, such as wood or gypsum board. For a single-layer construction, transmission loss at lower frequencies varies as $20 \log_{10} W$, where W is the surface mass per unit area. In general, doubling the thickness, and thus the mass, of a given material increases the low frequency TL by 6 decibels. At higher frequencies other factors relating to stiffness and damping come into play, and the relationship is no longer valid. For most practical purposes, the low frequency performance of a simple homogenous material is determined by mass or limp wall law, and the TL is determined by

$$TL_f = 20 \log_{10}(W/17) + 20 \log_{10}(f/63)$$

where W is the surface mass in kg/m² and f is the frequency in Hz. At some higher frequency the stiffness of the material causes the speed of flexural waves to match the speed of sound waves in air (5). In this frequency region the sound-isolating performance falls below the mass law TL . Above this region it increases but does not reach mass law performance again. The extent of the reduction depends on the internal damping of the material. Limp

materials, such as lead, have better sound-isolating performance for the same mass than stiffer materials, such as plaster. Table 4 provides the sound-transmission loss and STC ratings of some common materials.

Table 4. Sound-Transmission Loss of Some Common Materials

Material	Octave band center frequency, Hz						STC
	125	250	500	1000	2000	4000	
22 ga steel	13	17	23	28	34	39	27
1.3-cm gypsum board	15	20	24	30	31	27	28
6.3-mm acrylic plastic	15	18	22	28	32	35	27
1.0-cm plywood	14	18	22	20	21	26	22
1.6-mm lead foil	9	15	21	27	33	39	24

Uses. Sound-isolating constructions are needed around many types of rooms in buildings to reduce transmission of intrusive noise from exterior and interior noise sources and to provide acoustical privacy between rooms. Frequently the required degree of isolation can be provided using standard construction techniques. Music buildings, studios, performance facilities, and other acoustically critical spaces usually require special sound-isolating constructions to provide high degrees of sound isolation from exterior noises and between rooms. Increases in road and air traffic mean higher levels of environmental noise in many areas, and sound-proofing of residential properties and schools in the vicinity of highways and airports is being carried out in many locations in the United States. Within buildings sound-isolating constructions also are used to enclose noise-producing air-handling units and other mechanical equipment.

Sound-Isolating Constructions. Although some materials are used alone in single-layer constructions for sound-isolating purposes, most sound-isolating constructions contain two or more parts, frequently separated by an airspace.

Double-Wall Constructions. Significant improvements in *TL* can be achieved by using constructions consisting of two independent parts separated by an airspace (5,7). Double-glazed windows are one example of double-wall construction. Other common constructions of this type are steel stud partitions with gypsum board on both sides. Although the two sides are connected by the studs, the studs are relatively flexible and transmit a smaller amount of energy between the gypsum board faces than would be transmitted by a rigid connection; thus the acoustical performance of steel stud and gypsum board constructions approximates double-wall performance. Standard wood stud construction does not behave as a double wall because the stiffness of the studs provides rigid bridging between the gypsum board faces. Double-wall constructions have relatively poor performance at a specific low frequency, where a resonance of the two faces on the intervening air spring occurs, but above this frequency the *TL* increases rapidly. This double-wall resonance varies as the square root of the mass of the faces and the airspace separating them, so heavier constructions with large airspaces are more effective than lightweight ones with small airspaces, as illustrated in Figure 5 by the two double windows with different airspaces and glass thicknesses. The acoustical performance of many double-wall constructions can be improved by adding sound-absorbing material, such as fiber glass or mineral wool, in the cavity between the two faces. The increase in *TL* caused by the addition of fiber glass between the studs in a typical steel stud and gypsum board partition is shown in Figure 6.

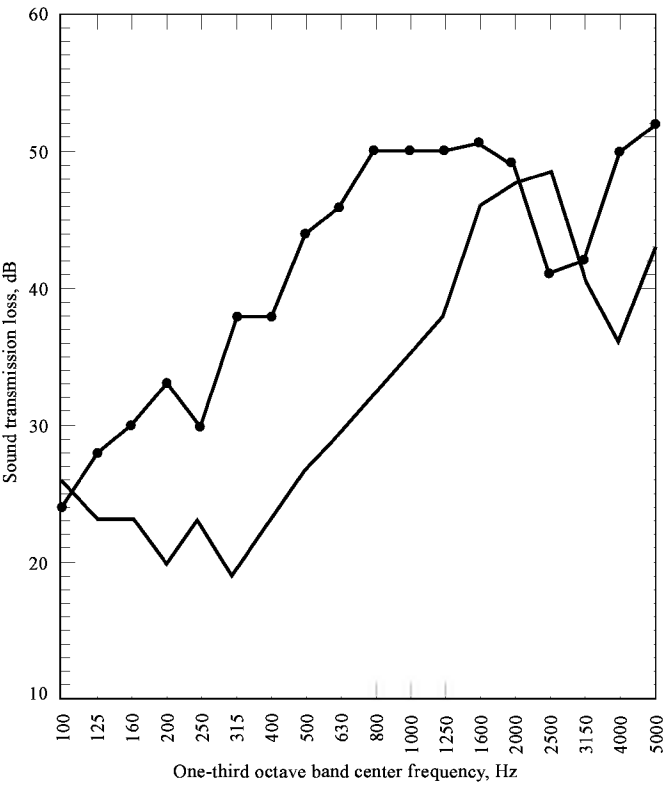


Fig. 5. Transmission loss of two types of double glass where (—) is 0.32 cm (1 8 in.) double glass and 0.95 cm (3 8 in.) air, and (—●—) is 0.48 cm ($\frac{1}{16}$ in.) double glass and 10 cm (4 in.) air.

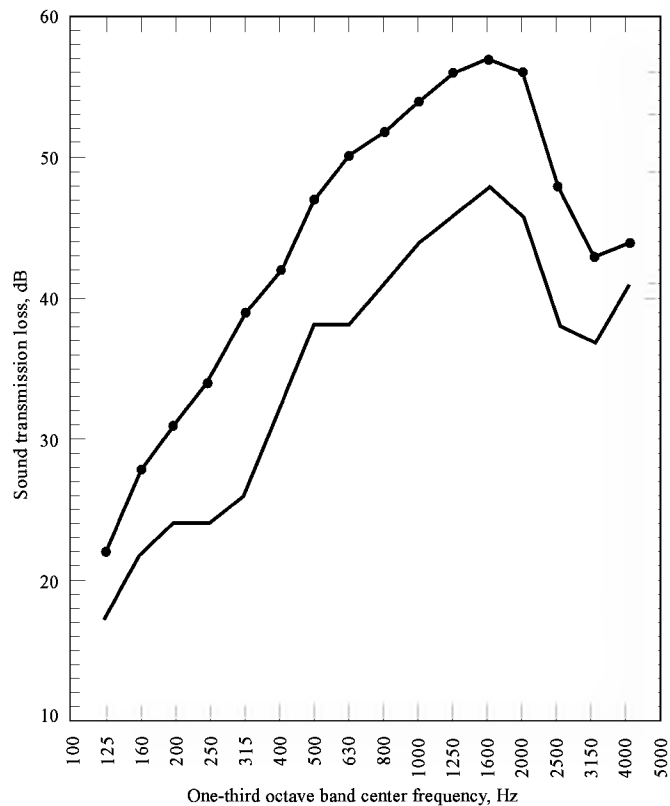


Fig. 6. Effect of fiber glass on transmission loss of steel stud partition where (—) is 6.35 cm (2S in.) steel stud and 1.27 cm (S in.) gypsum board, and (—■—) is 6.35 cm (2S in.) steel stud with 5 cm (2 in.) fiber glass.

STC ratings for wall constructions vary from about STC-15 for simple lightweight constructions to as high as STC-80 for heavy complex constructions. Ratings for some wall constructions are indicated in Table 5.

Table 5. STC Ratings of Some Common Building Constructions

Construction	STC
wood stud	
with 1.6-cm gypsum board both sides	35
with fiber glass insulation	38
double row of wood studs	
with 1.6-cm gypsum board on outside faces	45
with fiber glass insulation	56
9.1-cm steel stud	
with 1.6-cm gypsum board both sides	39
with fiber glass insulation	47
29-cm lightweight concrete block	47
29-cm dense concrete block	52
29-cm poured concrete	58
6.3-cm plate glass	31
solid wood door	
normally hung	15
fully gasketed	29

Impact-Isolating Constructions. Adequate impact sound isolation is difficult to achieve when hard materials, such as terrazzo, quarry tile, vinyl tile, hardwood, etc, are used on floors in multistory buildings. Complex constructions incorporating resiliently supported floors and or ceilings are required to reduce impact noise transmission when hard flooring materials are used. Carpeting can significantly reduce impact noise transmission, especially when installed over resilient padding. Because of the effectiveness of carpets and pads, many condominium associations require the owners to carpet significant portions of the floors in upper story units. Impact isolating class (IIC) and STC ratings of some floor ceiling constructions are provided in Table 6.

Table 6. Acoustical Performance of Floor/Ceiling Constructions

Construction	IIC	STC
20-cm reinforced concrete slab		
with marble floor	40	40
with hardwood floor on wood sleepers	45	44
with carpet and pad	70	40
5 × 25-cm wood joists, 41 cm on centers		
with oak flooring on plywood subfloor, gypsum board ceiling	35	39
with gypsum board on resilient channels	50	50
with carpet and pad	60	40

Vibration Isolation

Reciprocating, rotating, and rolling equipment not only generate noise, they also can transmit vibrational energy into supporting structures. These structures can transmit vibrations and or radiate unwanted sound into rooms or other occupied spaces where it can interfere with activities or cause annoyance. Vibration isolation devices reduce the transmission of vibrational energy between a supported vibrating object and the supporting structure.

Units. The performance of a vibration isolator is characterized by its transmissibility, defined as the ratio of the force transmitted to the supporting side of the isolator compared with the driving force acting on the vibrating side of the isolator (5,6):

$$\text{transmissibility} = \frac{\text{output force}}{\text{input force}}$$

The transmissibility of an isolator varies with frequency and is a function of the natural frequency (f_n) of the isolator and its internal damping. Figure 7 shows the transmissibility for a family of simple isolators whose fundamental frequency can be represented as follows, where k is the stiffness of the isolator, N/m; and m is the supported mass, kg. Figure 7 shows that an isolator acts as an amplifier at its natural frequency, with the output force being greater than the input force. Vibration isolation only occurs above a frequency of about $\sqrt{2}$ times the natural frequency of the isolator.

$$f_n = \frac{1}{2\pi} \sqrt{\frac{k}{m}}$$

Damping reduces the transmissibility at the natural frequency, but increases the transmissibility at higher frequencies. The natural frequency of isolators made from most materials also can be expressed as a function of the static deflection of the isolator due to the load imposed by the supported equipment; that is, $f_n = \frac{1}{5\sqrt{\delta}}$ where δ is the static deflection of the isolator, cm (5).

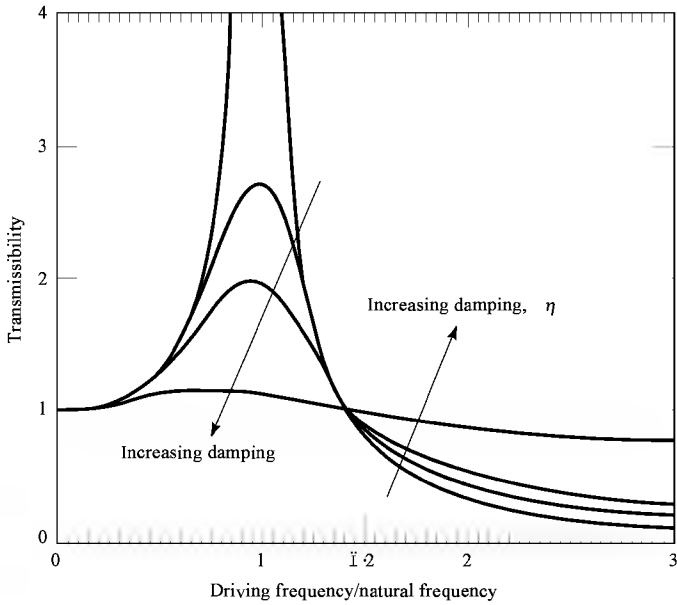


Fig. 7. Transmissibility as a function of frequency ratio for single-degree-of-freedom isolators with different degrees of internal damping.

Another measure of vibration isolation is isolation efficiency, which is one minus transmissibility and is usually defined as the percent of force transmitted through the isolator. Thus an isolator with a transmissibility of 0.75 has an isolation efficiency of 25%. A third measure of vibration isolation is insertion loss, which is the difference between the transmitted vibration with the isolators in place and with no isolators.

Test Methods. There is no standard test method for measuring transmissibility or isolation efficiency of vibration isolation devices. The most common procedure is to measure the vibration transmitted to the supporting structure with the isolators in place and with the equipment supported on rigid blocking. From these measurements the insertion loss in dB is determined by the following where L_i is the transmitted vibration with isolators in place and L_m is the transmitted vibration with rigid supports.

$$IL = 10 \log_{10} (L_i / L_m)$$

Materials. Materials commonly used in vibration isolators include steel in the form of springs, and elastomers (neoprene, natural rubber, glass fiber, cork, and felt) in the form of cubes and pads. The low frequency performance of steel spring isolators is superior to that of elastomeric isolators. They can be readily and repeatedly manufactured with predictable characteristics, and with the proper preparation they can be used in severe chemical and physical environments. A disadvantage of steel springs is that they tend to transmit high frequency vibrational energy. For this reason they are frequently used in combination with elastomeric elements, which are more effective in reducing high frequency vibrational energy.

Elastomeric materials, which provide relatively low practical static deflections and have relatively high natural frequencies, are used only to isolate higher frequencies. The volume compressibility of elastomeric materials is relatively low, therefore the shape of the elastomeric isolator must be taken into account, and space must be provided for lateral expansion. Because of their inherent resistance to chemical and environmental deterioration, neoprene and other synthetic materials often can be used in severe environments where natural materials would deteriorate.

Uses. In architectural and industrial applications vibrational isolators are used to reduce transmission of vibration into building structures from rotating or reciprocating machinery, such as ventilating fans, pumps, chillers, industrial machinery, and the piping and ductwork connected to this equipment (6). Vibration isolators also can be used to isolate vibration-sensitive equipment or noise-sensitive areas from sources of vibration. Examples are special pneumatic isolators to protect electron microscopes, and isolators used to support floating concrete floors in recording studios. Transportation-related applications include isolators used to reduce transmission of vibration from automobile and truck motors and exhaust systems into vehicle frames and bodies; from rapid transit and railroad steel rails into concrete inverts, bridges, and other supporting structures; and numerous other applications. Isolators also are used to minimize the vibration generated by fans and motors in various appliances. In many cases reducing transmission of vibration from the vibration-producing elements into structures having greater radiating efficiency can significantly reduce radiated noise.

Products. Vibration isolators typically are selected to have a static deflection, under load, that yields a natural frequency no more than one-third the lowest driving frequency that must be isolated (see Fig. 7). The supporting structure must have sufficient stiffness so it does not deflect under the load of the supported equipment by more than one-tenth the deflection of the isolator itself (6). In addition to static deflection requirements, vibration isolators are selected for a particular application according to their ability to carry an imposed load, and to withstand the environment in which they are used (extreme temperatures, chemical exposure, etc).

Commercially available vibration isolators include single and multiple coil springs with mounting bases and connectors for HVAC fans and other equipment. Frequently the springs are in series with neoprene or rubber elements. Spring hangers, neoprene hangers, and combinations of the two also are available for suspending vibrating equipment. Ribbed or waffled neoprene isolators typically are used to isolate equipment such as electrical transformers, which produce vibrational energy only at higher frequencies. At the other end of the spectrum are pneumatic isolators (air springs) consisting of inflated air bladders of neoprene or rubber with one or more tuned air chambers. They are used to isolate very low frequency vibrations. Several types of commercially available spring and neoprene vibration isolators are illustrated in Figure 8.

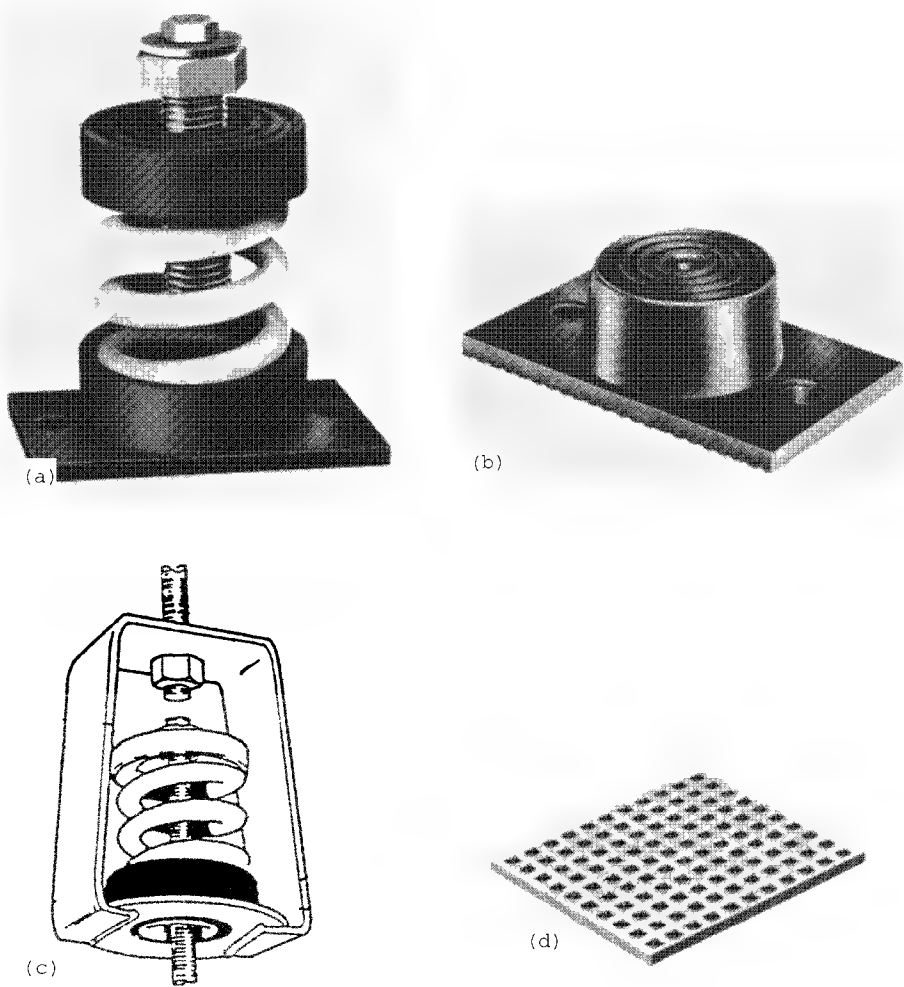


Fig. 8. Vibration isolators: (a) single-spring mount with base plate; (b) neoprene mount; (c) spring and neoprene hanger; and (d) neoprene waffle pad. Courtesy of Mason Industries, Inc.

Effective vibration isolation requires that there be no rigid connections between the isolated object and the supporting structure and other surrounding objects, because such connections short-circuit the isolators and reduce their effectiveness. An example of a short circuit that frequently is encountered in buildings is rigid electrical conduit connecting an isolated machine to the building structure from which it is being isolated. Such connections should be made with slack loops of flexible conduit or with special flexible electrical connectors. In some critical installations the conduits also should be vibration isolated.

Vibration Damping

Vibration damping is a process that reduces the vibrational energy in a system by converting some of the energy into heat. All materials and systems have some inherent damping, just as all materials absorb some sound, although in both cases the amounts can be very small. Damping is a highly complex phenomenon and there are many damping mechanisms, including interface friction, fluid viscosity, turbulence, acoustic radiation, eddy currents, and magnetic and mechanical hysteresis. Mechanical hysteresis is the only damping process that depends on internal friction within a material and is, therefore, also known as material damping.

Units. Two measures that are commonly used to define material damping are the loss factor (η) and amplification at resonance (Q), both of which are dimensionless: $\eta = D / 2\pi W = 1 / Q$ where D is the energy dissipated per cycle of vibration and W is the average total energy of the vibrating system. These relationships become more complex and less useful for highly damped systems. See References 5 and 7 for more extensive treatment of damping measures and treatments.

Test Methods. There are no national standards for the measurement of vibration damping. The most useful and convenient technique is to measure the reverberation time or decay rate of a panel or bar. The sample is vibrated by noise from a transducer, the noise is abruptly terminated, and the decaying vibrations are measured using an accelerometer to determine the decay rate Δ_t . The loss factor η is computed by $\eta = \Delta_t / 27.3 f_n$ where Δ_t is the decay rate, dB/s; and f_n is the natural frequency of the sample, Hz (5). One significant difficulty encountered in making these measurements is how to prevent excessive energy dissipation by the supporting system, the transducer, the accelerometer, and related cables. The sample may be suspended from long strings; the transducer should not contact the panel; the accelerometer should be as low in mass as possible; and the cables should be thin and flexible.

Materials. All materials have some inherent internal friction, or material damping. Most metals have relatively little damping, and conversely, high amplification at resonance. Strike a bronze bell and it will ring at its resonant frequency for an extended period. It has a low η ($< 10^{-3}$) and a high Q ($> 10^3$). Rubbery and soft materials have a higher η and lower Q . If a bell were made of acrylic plastic it would hardly ring at all when struck. Acrylic has a much higher η (3×10^{-3}) and lower Q (33) than that of bronze.

The loss factor (η) for most rigid materials such as metals, concrete, plywood, glass, etc, ranges from about 10^{-4} to slightly more than 10^{-2} . The loss factors for these materials do not vary much with frequency or temperature, and they are not high enough for these materials to be used for damping purposes. The loss factors for viscoelastic materials are orders of magnitude higher than for rigid materials; as a result, these materials are widely used for damping treatments. The maximum loss factors for these materials at room temperatures range from about 0.2 to about 5.0, but they vary widely with temperature and frequency. Because of these variations, viscoelastic materials intended for use as damping treatments must be selected to suit the frequency and temperature ranges of concern.

Uses. A damping treatment is a material or combination of materials applied to a metal panel or other structural element to increase its damping. The purpose of increasing the damping may be to reduce the vibration of the element at its resonant frequency, or it may be to attenuate flexural wave propagation along an extended structure, thereby increasing the sound-transmission loss of the structure. In both cases effective noise control is achieved only in very narrow frequency ranges; therefore caution should be exercised when using damping as a means of noise control.

Reduce Resonant Vibration. Metal structures are induced to vibrate at their natural frequencies when driven mechanically by attachment to some other vibrating structure, by impact of solid objects, or by turbulent impingement of a fluid (including air). Examples are stainless steel sinks driven

by garbage disposals; dishwasher cabinets impacted by water sprays; trash chutes and bins impacted by cans and bottles; tumbling bins, conveyors, and vibratory feeders impacted by small parts; and other devices that are periodically or continuously impacted by hard objects or attached to vibrating machinery. Damping treatments often can provide a considerable amount of noise reduction at the natural frequency of this type of sheet metal structure when applied to its radiating surfaces, but the treatment has no significant effect at nonresonant frequencies.

Increase Sound-Transmission Loss. The only significant increases in sound-transmission loss that can be achieved by the application of damping treatments to a panel occur at and above the critical frequency, which is the frequency at which the speed of bending wave propagation in the panel matches the speed of sound in air. Application of damping treatment to 16 ga metal panel can improve the *TL* at frequencies of about 2000 Hz and above. This may or may not be helpful, depending on the application of the panel.

Another practical application of damping to increase sound-transmission loss is the fabrication of acoustical glass by laminating a soft vinyl interlayer between two sheets of glass. This lamination improves the *TL* in two ways: first, by raising the critical frequency above that of the same thickness of monolithic glass; and second, by providing damping, which reduces flexural wave propagation in the glass.

Products. Damping treatments are available from many manufacturers in sheet form, as tapes for adhering to a surface, and in bulk form for spraying or troweling onto a surface. Laminated glass is available from many glass suppliers (see LAMINATED MATERIALS, GLASS).

Extensional or Free-Layer Treatments. These are viscoelastic damping treatments that are applied directly to a surface in a variety of ways. A free viscoelastic layer stores and dissipates energy primarily as a result of the stretching and compression of the layer caused by bending. The damping provided by this type of treatment increases roughly as the square of the layer's thickness until it reaches a thickness about three times that of the surface to which it is applied. Above this thickness the increase is less rapid. Free-layer damping treatments tend to use viscoelastic layers that are between one-half and two times as thick as the underlying structure. They must be continuous and well-bonded to the structure. Some sheet damping products are available with a self-adhesive backing and others are applied using a thin layer of epoxy or some other rigid adhesive.

Constrained-Layer Treatments. Constrained-layer damping treatments consist of a thin layer (μm) of viscoelastic material sandwiched between a base material and an outer constraining layer of sheet metal or other structural material. Some of these treatments are available with self-adhesives on both sides of the viscoelastic material and act as a bonding agent between the base and constraining layers; others have the constraining layer already bonded to the inner layer so they need only be applied to the base material.

Sound-Absorptive Blankets. Sound-absorptive blankets of fiber glass or mineral wool are not usually considered damping materials, but when fastened to sheet metal machine enclosures they can provide some useful damping in addition to sound absorption.

BIBLIOGRAPHY

"Insulation, Acoustic" in *ECT* 3rd ed., Vol. 13, pp. 513–534, by C. R. Jokel, Bolt, Barenek & Newman.

1. *Test Method for Sound Absorption and Sound Absorption Coefficients by the Reverberation Room Method, ASTM C423-90a*, ASTM, Philadelphia, Pa., 1990.
2. *Standard Practices for Mounting Test Specimens During Sound Absorption Tests, ASTM E795-92*, ASTM, Philadelphia, Pa., 1992.
3. *Test Method for Impedance and Absorption of Acoustical Materials by the Impedance Tube Method, ASTM C384-90a*, ASTM, Philadelphia, Pa., 1990.
4. *Method for Impedance and Absorption of Acoustical Materials Using a Tube, Two Microphones, and a Digital Frequency Analysis System, ASTM E1050-90*, ASTM, Philadelphia, Pa., 1990.
5. L. L. Beranek, ed., *Noise and Vibration Control*, McGraw-Hill Book Co., New York, 1971.
6. *1991 ASHRAE Handbook, Heating, Ventilating, and Air Conditioning Applications*, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, Ga., 1991, Chapt. 42.
7. L. L. Beranek and I. L. Ver, eds., *Noise and Vibration Control Engineering*, John Wiley & Sons, Inc., New York, 1992.
8. *Classification for Rating Sound Insulation, ASTM E413-87*, ASTM, Philadelphia, Pa., 1987.
9. *Standard Test Method for Laboratory Measurement of Impact Sound Transmission Through Floor–Ceiling Assemblies Using the Tapping Machine, ASTM E492-90*, ASTM, Philadelphia, Pa., 1990.
10. R. D. Berendt and G. E. Winzer, *A Guide to Airborne, Impact, and Structure Borne Noise Control in Multifamily Dwellings*, U.S. Department of Housing and Urban Development, Washington, D.C., 1963.
11. *Standard Test Method for Laboratory Measurement of Airborne Sound Transmission Loss of Building Partitions, ASTM Designation E90-90*, ASTM, Philadelphia, Pa., 1990.
12. *Standard Test Method for Measurement of Airborne Sound Insulation in Buildings, ASTM E336-90*, ASTM, Philadelphia, Pa., 1990.

General References

Catalog of STC and IIC Ratings for Wall and Floor/ Ceiling Assemblies, Office of Noise Control, California Department of Health Services, Berkeley, Calif., 1984.

R. S. Jones, *Noise and Vibration Control in Buildings*, McGraw-Hill Book Co., Inc., New York, 1984.

D. A. Bies and C. H. Hansen, *Engineering Noise Control Theory and Practice*, Unwin Hyman Ltd., London, 1988.

C. M. Harris, ed., *Handbook of Acoustical Measurement and Noise Control*, 3rd ed., McGraw-Hill Book Co., Inc., New York, 1991.

M. D. Egan, *Architectural Acoustics*, McGraw-Hill Book Co., Inc., New York, 1988.

Compendium of Materials for Noise Control, HEW (NIOSH) publication 75-165, HEW, Washington, D.C., June 1975.

C. M. Harris, ed., *Noise Control in Buildings*, McGraw-Hill Book Co., Inc., New York, 1994.

Parker W. Hirtle
Carol E. Parssinen
Acentech Incorporated

INSULATION, ELECTRIC

The U.S. Department of Commerce Bureau of Census reports that in 1990 the total value of insulated wire and cable shipments was more than \$10.5 billion. These shipments have grown more than 165% compared to 1983 and more than 225% compared to 1977.

Relative sizes of the principal market segments of insulated wires and cables have changed dramatically since 1977: the electronic wires segment has grown by 430%, power wires and cables have grown by 288%, and building wires and cables by 283%. Some other segments have shown smaller increases and even decreases; most notably, in 1990 telephone cables were only 93% of the 1977 shipments. Table 1 compares the values of various segments of insulated wire and cable sales in 1977 and 1990 (1,2).

Table 1. Insulated Wires and Cables Sales

Market	1977		1990	
	Sales, 10 \$	% of total	Sales, 10 \$	% of total
electronic wires	474.1	10.0	2,039.2	19.0
telephone cables	1,619.8	34.2	1,514.8	14.2
power cables	695.4	14.7	2,005.4	18.8
control and signal wires	125.2	2.7	251.0	2.4
building wires	875.4	18.5	2,477.8	23.2
automotive wires	191.6	4.0	366.4	3.4
appliance wires	169.7	3.6	296.7	2.8
other equipment wires	293.5	6.2	436.1	4.1
line and extension cord	290.5	6.1	534.6	5.0
fiber optics cables	negligible		758.1	7.1
Total insulated wire	4,735.2	100	10,680.1	100

Almost all of the industry segments mentioned represent wires and cables that conduct electricity using currents of relatively high voltage and amperage but low frequency for power cables, or currents of high frequencies but low voltage and amperage for telephone and electronic wires. One segment of the insulated wire industry that was not mentioned in the 1977 Bureau of Census report but which has grown dramatically is fiber optic cables. It has been predicted that by the year 2000 fiber optic cables will become the largest segment of the industry, mostly at the expense of the telephone and electronic communication wires. These cables do not conduct electricity, but rather use light as the vehicle for communicating data (see FIBER OPTICS).

Each segment of the insulated wire and cable industry has its own set of standards, and cables are built to conform to specifications provided by a large variety of technical associations such as The Institute of Electrical & Electronic Engineers (IEEE), The Insulated Cable Engineers Association, (ICEA), National Electrical Manufacturers Association (NEMA), Underwriters Laboratories (UL), Rural Electrification Administration of the U.S. Department of Agriculture (REA), Association of Edison Illumination Companies (AEIC), Military Specifications of the Department of Defense (MIL), American Society for Testing and Materials (ASTM), National Electrical Code (NEC), etc.

Designs and Materials

Data Communication Wires. Electronic cables such as data communication wires employ three basic designs: coaxial, twisted pair, and fiber optics (3,4) (Fig. 1). Coaxial cables are so named because the axis of curvature of its outer conductor is concentric to its inner central wire. The metal braiding wrapped around the insulated center wire acts as the return current conductor in addition to shielding the wire from various interferences.

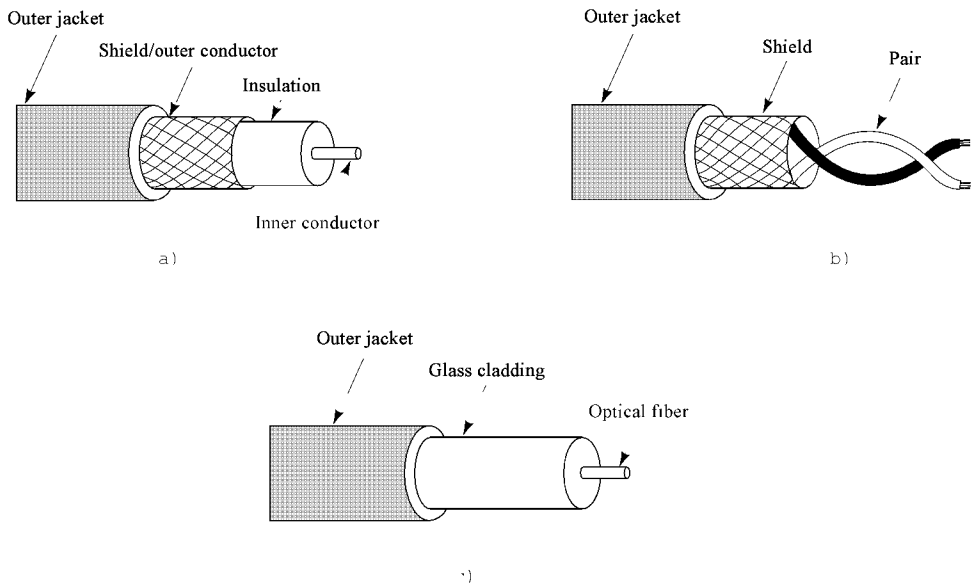


Fig. 1. Cable designs: (a) coaxial cable; (b) twisted pair cable can be unshielded, as in regular telephone wiring, or shielded (as shown here) with braiding or foil; (c) fiber optics cable.

The twists of twisted pair cable act as a shield against radio frequency interference (RFI), and electromagnetic interference (EMI), and against the cross talk interference that a wire exerts on nearby wires; the more twist the less interference. Telephone wires can use large numbers of pairs. In most cases the pairs are not shielded with braiding or foil, as shown in Figure 1b for data communication wire. Data communication wires work at very high

current frequency (GHz), and can transmit a very large quantity of digital data, as opposed to modulated currents at low frequency that convey lower amounts of data used by the telephone wires.

Fiber optic transmission works differently from copper cable transmission. Instead of metal wire conducting an electrical charge, hair-thin glass or plastic fibers conduct light that is sent by rapid flashing on and off (digitally). The light can travel in more than one beam (multimode) or in one monomode beam, by bouncing off the inner walls of the hollow fiber (3). The coating layer cladding helps reflect light down the fiber.

Cables are available in a variety of constructions and materials, in order to meet the requirements of industry specifications and the physical environment. For indoor usage, such as for Local Area Networks (LAN), the codes require that the cables should pass very strict fire and smoke release specifications. In these cases, highly flame retardant and low smoke materials are used, based on halogenated polymers such as fluorinated ethylene–propylene polymers (like PTFE or FEP) or poly(vinyl chloride) (PVC). For outdoor usage, where fire retardancy is not an issue, polyethylene can be used at a lower cost.

Building Wires. These wires conduct electricity at relatively low voltages (eg, 110 V and 220 V). Typically they contain a metallic conductor (copper or aluminum) that is insulated with polymeric compounds based on polyethylene or PVC which are applied over a conductor using an extruder.

Magnet Wires. These wires are used principally in the electrical and electronics industries for coils, inductors, transformers, armatures, solenoids, etc. Typically the manufacturing process takes the metallic conductors through a liquid bath (or sometimes a powdered fluidized bed) of varnishes and enamels based on special resins, such as polyesters, polyurethanes, poly(vinyl formal), or polyimides followed by a heating process that drives off the solvents and cures these resins in a hard nonmelting thermosetting layer. Magnet wires may be classified according to NEMA thermal ratings based on extrapolations of their thermal lives measured in laboratory conditions.

Specialty Wires. Several categories of specialty wires employ special designs and materials, custom made to fit particular applications and or specifications (5–8) (Fig. 2).

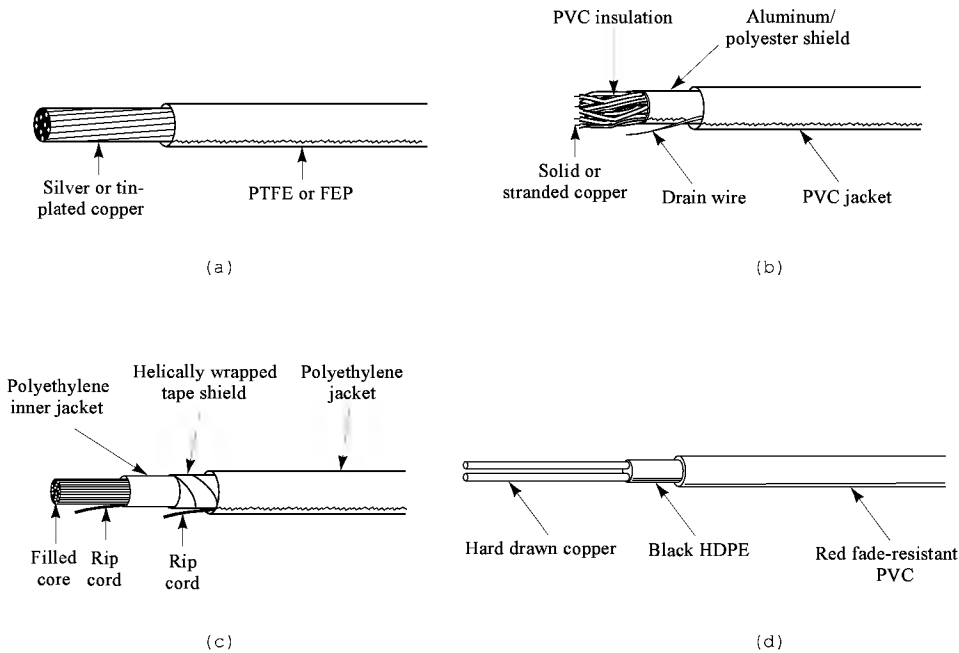


Fig. 2. Specialty wires: (a) appliance wires; (b) instrumentation wires; (c) distribution wires; and (d) aerial self-supporting wires.

Appliance wires require a higher temperature rating (105°C or higher). Therefore, the insulation is made of fluorinated thermoplastics, such as poly-tetrafluoroethylene (PTFE) or fluorinated ethylene–propylene (FEP).

Cross-linked polyethylene-based compounds that contain flame-retardant components and compounds based on PVC cross-linked by radiation have also received high temperature rating. They find use not only in appliance wires but also in manufacturing under-the-hood *automotive wires*.

Instrumentation wires contain multiple pairs of conductors, each insulated with flame-retardant PVC and with an overall flame-retardant PVC jacket (5). For *distribution wires*, polyethylene or ethylene–propylene rubber are the polymers of choice (Fig. 2c). A typical design for *aerial self-supporting wires*, that employs PE and PVC, is shown in Figure 2d.

Military Application and Aerospace Wires. Depending on the specific application, a variety of polymers can be considered: PVC, polyamides, PTFE, etc (Fig. 3). Navy shipboard specifications require cables with flame retardancy, low smoke emission during fire, and containing no halogen.

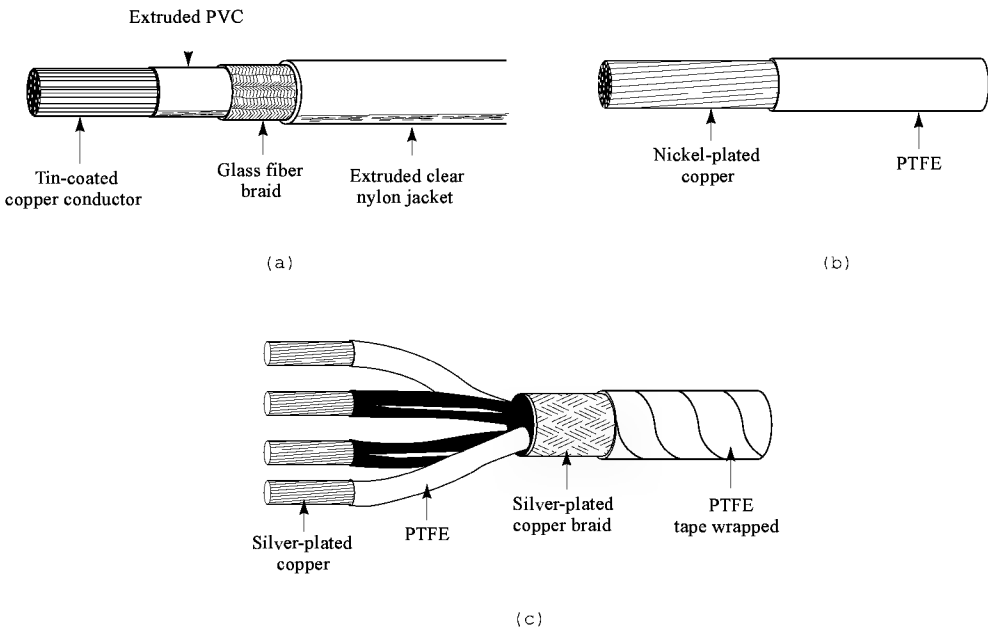


Fig. 3. Military application and aerospace wires.

Railroad/Transit Cables. These are single and multiconductor cables, rated 300 to 2000 V. The cables are designed for railroad and transit applications including vital circuits, track circuits, train control, third rail feeders, or apparatus wiring. Installation may be in wet as well as dry locations, in subway tunnels, or directly buried in the earth. Their insulation can be based on ethylene–propylene rubber (EPR) and is specially compounded to be flame retardant; the jacket can also be flame retardant with low smoke emission during fire. Specifications require that during fires the transit cables should exhibit low smoke emission, low toxicity, low corrosivity; some specifications do not allow the use of halogenated materials in cable composition. The issue of halogen-free cables has been under discussion in the 1990s, especially for cables placed in closed environments, such as underground public transportation (transit, railways), buildings which house large numbers of people (eg, department stores, hospitals, offices, hotels), buildings which house valuable installations (eg, telephone and computer centers, power stations, television and radio stations), and military installations (eg, Navy ships, submarines). When fires occur in enclosed spaces, the halogenated compounds can decompose and release toxic and corrosive chemicals such as hydrochloric acid, which are harmful to health and corrosive to important and expensive equipment.

Control and signal cables are made up of fine copper wire strands of plain electrolytic copper wire with PVC or EPR-based insulation and an outer jacket of special PVC or ethylene copolymers.

Electric Submersible Oil Well Pump Cable. These cables are rated up to 5 kV and are designed for highly corrosive oil wells that besides oil also contain brine and other harsh chemicals as well as gases under high pressure and high temperatures (6). Insulations can be based on polypropylene for low temperature wells or on ethylene–propylene rubber which is compounded with special ingredients in order to resist the environments of high temperature wells (Fig. 4).

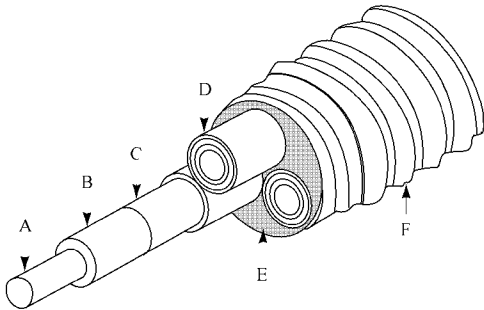


Fig. 4. Submersible oil well pump cable. A, Solid copper conductor; B, EPR-based insulation; C, chemical barrier; D, lead sheath; E, filler; F, galvanized steel armor.

Power Cables. These high voltage cables have the most complicated designs. Depending on the voltages used (138,000 + V), the power cables can contain many layers, each one made of specially developed materials, with very specific characteristics (9). Figure 5 shows a typical 5-35 kV distribution power cable (9), such as the URD (Underground Residential Distribution) power cable. Typical conductors are aluminum or copper, mostly stranded or solid. At special request filled strands based on organic compounds are used. The conductor shield is bonded to the insulation. Conventional conductor shields are semiconductive and contain an ethylene-based copolymer and large amounts of carbon black. Some companies promote stress-relieving layers based on high dielectric constant (high permittivity) materials. Conventional insulations may employ either cross-linked polyethylene (XLPE), tree retardant cross-linked polyethylene (TRXLPE), or EPR-based compounds. Conventional design for the insulation shield uses semiconductive compounds based on carbon black loaded ethylene-based copolymers that are thermosetting in nature and are bonded, but also strippable from the insulation layer. Concentric copper neutral wire is used for returned current. The jacketing layer can be based on thermoplastic polymers, such as polyethylene, PVC, or thermosetting compounds based on polymers like chloroprene, chlorosulfonated polyethylene, nitrile, chlorinated polyethylene, etc; it can be insulating or conductive. Some power cables have a metal shield at special request.

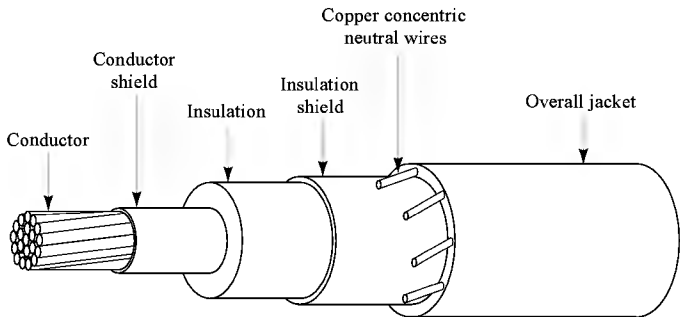


Fig. 5. Distribution power cable.

Most of the polymeric-based layers are applied using extrusion technology; the main equipment is the extruding coating line (see POLYMER PROCESSING).

Properties and Test Specifications

Each segment of the insulated wire and cable industry has its own set of standards, some of which are quite complicated because of requirements imposed by specific applications and or environments. The most complex specifications are typically imposed on power cables and telecommunication wires.

The most important electrical properties of insulation are dielectric strength, insulation resistance, dielectric constant, and power factor. Corona resistance, although not strictly an electrical property, is usually considered also (10).

Dielectric Strength. The dielectric strength of a material is the electric stress required to puncture a sample of known thickness and is expressed in terms of volts per thickness units, eg, V / μm. The dielectric strengt of an insulating material is influenced by the rate of rise of the applied voltage, and the total length of time the voltage is applied. A slow rate of rise usually causes the material to puncture at a lower voltage than does a rapid rate of rise. Similarly a material may withstand a relatively high voltage for a short time, but is punctured by prolonged exposure to a considerably lower voltage.

Dielectric strength is measured by determining the minimum voltage which will puncture a sample of known thickness placed between electrodes of specified size and shape. Because both the magnitude and duration of the applied voltage influence the results, this property can be measured in three ways.

In the most frequently used test the sample is placed between two electrodes and the voltage is increased from zero at a uniform rate until breakdown occurs. When an insulated wire is available, the voltage can be placed between the inner conductor and a conductive medium, such as an outside metallic shield or even water.

Another test consists of the application of a voltage starting at zero and increasing at a uniform rate up to a predetermined value. The voltage is held at this value for a specified time. The voltage and time vary with the type of product and with the kind of information desired. This test is useful for determining whether or not a given product or assembly has a sufficient high dielectric strength. It is nondestructive and the voltage applied is determined more by service conditions than by the actual dielectric strength of the insulating material. Since failures are caused by manufacturing defects, impurities, or damage, this test is used extensively for quality control.

A third test consists of an instantaneous application of the full test voltage; higher voltages are impressed on the insulation than in previous tests. A spark test is used to continuously detect faults in wire insulation during some stage of its manufacture. This test employs a chamber which contains either a bath of metallic spheres or a chain curtain suspended from the top of the chamber. The wire is run through the chamber and high voltage is applied between the beads or curtain and the insulated conductor. The voltage is held just under the maximum stress the insulation can withstand, so that any foreign material, thin spot, or other defect will cause a spark to pass through the insulation.

In actual practice, mechanical and electrical design factors usually require the cables to have layers of a certain thickness such that the electrical stress is far below the dielectric breakdown point.

Resistivity/Conductivity. The resistivity or specific resistance of a material is the electric resistance offered by an element of the material having unit length and unit cross-sectional area. The current intensity is proportional to the voltage across its path, and is inversely proportional to resistance. This relationship is expressed by Ohm's law, where I = current in amperes, E = potential in volts, and R = resistance in ohms.

$$I = \frac{E}{R}$$

The resistance of a segment of the path described above is proportional to its length, inversely proportional to its cross-sectional area, and proportional to a specific property of the material of the segment called resistivity or volume resistivity, ie:

$$R = \rho \frac{L}{A}$$

where R = resistance of the segment, L = length of the segment, A = cross-sectional area of the segment, and ρ = resistivity of the material of the segment. The resistivity of a material is therefore

$$\rho = \frac{RA}{L}$$

The reciprocal of resistivity is conductivity.

There is no perfect conductor, nor is there a perfect insulator, hence every material has some value of resistivity. The range of resistivity values between good conductors and good insulators is tremendous. A conductor such as copper has a resistivity of about $1.7 \times 10^{-6} \Omega\text{cm}$ as compared with the resistivity of an insulator such as polyethylene, which is $\sim 10^{17} \Omega\text{cm}$ or more.

For flat samples such as press cured slabs the resistivity may be computed from the following formula where ρ = resistivity, R = resistance, A = area of the sample (the effective area of the smaller electrode if two electrodes of different sizes are used), and t = thickness of the sample.

$$\rho = \frac{RA}{t}$$

For wire insulation, the relationship between the resistance of the sample and the resistivity of the insulation material is expressed by the following equatisend on, where R = resistance, ρ = resistivity, L = length of the sample, r_1 = radius of the conductor, and r_2 = radius of the insulation.

$$R = \frac{\rho}{2\pi L} \log_e \frac{r_2}{r_1}$$

This formula may be rearranged for greater convenience, and at the same time D_2/D_1 may be substituted for r_2/r_1 where D_2 is the diameter of the insulation and D_1 is the diameter of the conductor. Also, it is more convenient to use common rather than natural logarithms.

$$\rho = \frac{2\pi RL}{2.3 \log_{10} \frac{D_2}{D_1}}$$

The method of measuring insulation resistance varies with each type of device or product. The insulation resistance of insulated wire is the resistance between the conductor and the outside of the insulation. When the insulation is covered by a metallic sheath or braid the measurement is made between the conductor and the sheath. Insulated wire with no sheath is usually immersed in water and the resistance measured between the conductor and the water after the wire has been immersed for a specified period of time.

For each specific application of a rubber compound as an insulating material, there is a minimum value of resistivity below which it does not function satisfactorily. In addition, insulating compounds are required to withstand the effect of water, moist atmosphere, or heat without their resistivity values falling below a satisfactory level. Insulation resistance measurements frequently serve as useful control tests to detect impurities and manufacturing defects in rubber products.

Dielectric Constant. Dielectric constant or specific inductive capacity (SIC) is both defined and measured by the ratio of the electric capacity of a condenser having that material as the dielectric to the capacity of the same condenser having air as the dielectric. The dielectric constant of vacuum is unity. Dry air has a constant slightly higher, but for most practical purposes it is considered as unity.

Two parallel plates of conducting material separated by an insulation material, called the dielectric, constitutes an electrical condenser. The two plates may be electrically charged by connecting them to a source of direct current potential. The amount of electrical energy that can be stored in this manner is called the capacitance of the condenser, and is a function of the voltage, area of the plates, thickness of the dielectric, and the characteristic property of the dielectric material called dielectric constant.

The capacitance of a condenser in terms of its physical dimensions and the dielectric constant of the insulation is given by the following equation, where C = capacitance in microfarads, K = dielectric constant of the insulation, A = area of plates in square centimeters, and t = thickness of the insulation in centimeters.

$$C = 0.088 \frac{KA}{t}$$

If an alternating current potentialis applied to an electrical condenser, each reversal of the potential results in a reversal of the charge stored in the condenser. There is, therefore, an alternating current apparently flowing through the condenser proportional to the capacitance of the condenser, hence proportional to the dielectric constant of the insulation material forming the dielectric of the condenser.

The dielectric constant of the insulation of a wire is measured by immersing a known length of wire in a conducting medium such as water or mercury. The dielectric constant vs capacitance relationship for a wire is given by the following formula, where C = capacitance, L = length of wire, K = dielectric constant of the insulation, r_2 = radius of insulation, and r_1 = radius of conductor. The most commonly used length of sample for this test is ~6 m (20 ft) immersed length.

$$C = \frac{LK}{18 \times 10^5 \log_e \frac{r_2}{r_1}}$$

Typical dielectric constant values for raw materials are 2.6–3.0 for natural rubber insulation, approximately 2.2 for polyethylene, and approximately

2.4 for ethylene–propylene rubber.

For most commercial voltages and frequencies used in power distribution, the capacitance effects are negligible. At relatively high voltages the current due to capacitance may reach sufficient value to affect the circuit, and insulation for such an application is designed for a moderately low dielectric constant.

The dielectric constant of a compound is increased by small amounts of absorbed water; hence wire insulation for communications generally must have a dielectric constant as stable as possible in the presence of water or moisture.

For telecommunication wires, where higher frequencies are used, there are some other critical properties that are related to the dielectric constant; for example, mutual capacitance, defined as the capacitance between two wires of a pair. In voice communication, mutual capacitance shifts the phase of the transmitted analogue signal. Since voice frequencies vary over a narrow range, phase shifts are usually not objectionable. In high frequency digital transmission, however, metal capacitance rounds or distorts the square wave shape of the signal, causing error in data transmission. The larger capacitance, the higher the distortion and error rate.

For coaxial cables, the following electrical properties related to the dielectric constant of the core material and the dimensions determine the quality of the signal: impedance, capacitance, attenuation, crosstalk, and time delay and velocity of propagation.

Impedance. Impedance defines the relationship of voltage and current in a coaxial cable. The electrical requirements of the hardware dictate the impedance values for the interconnecting cables. Most coaxial cables are designed to match the impedances required by electronic hardware.

Capacitance. This property is dependent on the dimensions of the inner and outer conductors and the dielectric constant of the core. Most computer systems have a maximum allowable capacitance for interconnecting cables. For these systems, the lower the capacitance of the core material, the longer the cable that can be used.

Attenuation. Attenuation refers to the reduction in amplitude or height of a transmitted signal. In voice communication, attenuation simply means that the conversation is not as loud. Attenuation of a digital signal reduces the height of the square wave so that the receiving equipment must be sensitive enough to distinguish the signal's on and off states and the difference between an adjacent signal. If the receiver has to look too closely, it can be deceived by noise pulses, causing errors in the data.

Because the FCC limits the strength of the transmitted signal, increasing the strength of the original signal is not an acceptable solution. Therefore, low attenuation is essential for high quality, error-free signal transmission, particularly over long cable runs.

Combined Effect of Capacitance and Attenuation. When capacitance is high, the signal never reaches the 1 state before it starts declining to 0 again. This yields a signal in which the 1 and 0 states are nearly indistinguishable by the receiver and an error results. Since capacitance and attenuation are always present in telephone cables, for error-free transmissions the communications wire must have the lowest capacitance and attenuation possible.

Crosstalk. This is a measure of the signal induced in a quiet pair by an excited pair. The excited signal could be voice, digital data, ringing, or noise. Crosstalk is expressed as a decibel (dB) loss, so the smaller the number, the less the crosstalk. Crosstalk becomes important when transmitting digital signals at high speeds.

The relationship of the dielectric constant of the cable insulation to crosstalk can be measured by testing two cables for crosstalk with the same dimension, but different insulation materials. The cable with the lower dielectric constant has less capacitance unbalance, thus resulting in lower crosstalk than the cable with the higher dielectric constant.

Time Delay and Velocity of Propagation. Time delay is directly proportional to the square root of the dielectric constant and describes the time that it takes for a signal to travel through a cable. The lower the dielectric constant, the less time required for a signal to travel through a cable.

Velocity of propagation is the speed of transmission in a cable as compared to the speed of transmission in air and is therefore expressed as a percentage. Since the velocity of propagation is inversely proportional to the square root of the dielectric constant of the core, a lower dielectric constant results in higher transmission speed (3).

Power Factor. The amount of energy given up by the condenser during discharge is measured. The power factor is the ratio of this loss to the energy required to charge the condenser and may be expressed as a decimal fraction or a percent of the charging energy. The equipment for measuring power factor is the same as for measuring dielectric constant, and usually the two are determined simultaneously.

The power factor of a sample is determined from the capacitance and resistance values by means of the following relationship, where P = power factor, G = conductance in mhos (reciprocal ohms), $W = 2\pi \times$ frequency, and C = capacitance.

$$P = \frac{G}{WC}$$

Typical power factors for an EPR-based compound employed for 5–35 kV power cable is approximately 0.03–0.05% when measured at room temperature and about 1.0–1.4% measured at 90°C.

Power factor, like the dielectric constant, is a property that represents a power loss that takes place when a wire insulation becomes the dielectric of a condenser because of a surrounding sheath or other conducting medium.

Power factor losses under certain conditions cause a temperature rise in the insulation that may result in failure or reduced life of the insulation. In communication wiring the power factor of the insulation plays an important role. Here the actual power loss can represent an appreciable portion of the total energy in the circuit. In addition, this loss disturbs the circuit characteristics of the equipment at both ends of the line.

Corona Resistance. Corona resistance is the ability of material to withstand the effect of electrical discharge. Corona discharge is a flow of electrical energy from a conductor at high potential to the surrounding air. If the cable has an insulating covering, the corona discharge takes place at the outer surface of the insulation. If there are voids or air spaces between the conductor and its insulation, corona discharge (sometimes named partial discharge) will probably take place at these points. The discharge is accompanied by a faint glow and a noise, and can convert oxygen to ozone and ionize gases.

The insulation on the conductor is therefore exposed to a considerable concentration of ozone and subjected to chemical reactions and mechanical erosion from the impingement of ions. This causes deleterious effects and shortens the life of the cable.

There are several methods to determine and compare the resistance to partial discharges. Some tests are done on finished cables, such as the U-bend test, and others are done on laboratory samples molded from the insulation, that are subjected to partial discharges created by sharp objects, such as needles under high voltages. The tests compare either the energy required or the length of time required to erode or fail (short circuit) samples of similar thickness.

Electrical and Water Treeing

Treeing is an electrical prebreakdown phenomenon. This type of damage progresses through a dielectric section under electrical stress so that, if visible, its path looks something like a tree. Treeing can occur and progress slowly by periodic partial discharge, it may occur slowly in the presence of moisture without partial discharge, or it may happen rapidly as the result of an impulse voltage. Although generally associated with a-c or impulse voltages, treeing has been observed with high d-c voltage stresses in wet experimental conditions. Treeing may or may not be followed by complete electrical breakdown of the dielectric section in which it occurs. In solid organic dielectrics it is the most likely mechanism of electrical failures which do not occur catastrophically, but rather appear to be the result of a more lengthy process.

Generally, trees occur under the relatively high voltages associated with power cables (11–13). Trees can be classified in three classes: electrical, water, and electrochemical.

Electrical trees consist of visible permanent hollow channels, resulting from decomposition of the material, and show up clearly in polyethylene and other translucent solid dielectrics when examined with an optical microscope. Fresh, unstained water trees appear diffuse and temporary. Water trees consist of very fine paths along which moisture has penetrated under the action of a voltage gradient. Considerable force is required to effect this

penetration which starts at a surface imperfection or stress concentration and must rupture but not decompose the internal structure as it progresses. When the voltage force and source of water are removed, most of the injected water diffuses away and evaporates, and the tree disappears. This disappearance indicates that channels or paths close up, because if they did not, their appearance would be enhanced rather than diminished when the water is replaced by air which has a greater refractive index difference with respect to polyethylene.

Electrical and water trees can grow from the interface of electrode and insulation into the insulation or they can grow from internal voids and contaminant particles radially outward, parallel to the field, and toward the electrodes. These latter are called bow tie trees. Trees which start their growth at surfaces with an unlimited supply of air or water can grow completely through a dielectric section to bridge the electrodes. These are called vented trees. Trees which start at an internal void or inclusion are called nonvented trees and rarely grow very large.

Electrochemical treeing is applied in those cases of water treeing in which the water contains solute ions which move under the action of an electric field and are detected within the insulation layer, or at an electrode surface after having passed through the insulation. They are not encountered as often as the first two classes, for example, trees formed in a cable exposed to a hydrogen sulfide environment called sulfide trees.

Test Methods for Electrical and Water Treeing.

Laboratory Samples. In order to test resistance against electrical treeing, the concept of the standard defect is used in the needle test and modifications thereof. Since trees initiate and grow at sites of stress concentration rather than in perfectly uniform fields, the needle test provides a reproducible and highly divergent electrical field when the specimens are prepared with precision. In this test, the needles have very sharp and well-defined tips and are inserted in sample materials at defined depths; the samples are electrically stressed under certain voltages for periods of time. After stressing, the specimens are carefully examined with a 100× microscope to determine evidence of trees at the needle tip. There are also several laboratory methods to test resistance to water treeing formation. For opaque specimens, such as filled EPR-based compounds, there are special techniques that include special staining chemicals that color electrical and even water tree paths and make them visible (11).

Tests on Cable Constructions. The Association of Edison Illumination Companies (AEIC) has approved an accelerated cable life test in which typical underground distribution power cables can be statistically compared based on their resistance to water treeing (number of days to fail). The comparison can be made by varying the type of insulation and or other cable layers in an environment that contains hot water (90°C) under 8 V μ (200 V mil) voltage stresses (four times the typical power cables operating voltages).

Physical, Mechanical, and Environmental Tests

Typical standard tests performed on insulation and or jacket compounds measure tensile strength, ultimate elongation, modulus, set, tear, heat distortion, heat shock, cold bend and low temperature brittleness, abrasion resistance, and shear resistance. Depending on the environment in which the cable operates, the following tests may be done: resistance to oil or other chemicals, including water absorption; air aging resistance, measured at various temperatures either as percent retention of the sample initial physical properties or as the ultimate end life for sample to become brittle; oxygen and ozone resistance; radiation resistance when used in nuclear stations; flame resistance, measured as oxygen index or vertical or horizontal flame tests; smoke tests, using various equipment; and flame and smoke emission for the wires used indoors in the plenum areas are determined by the UL910 test.

Materials Used in Insulated Wires and Cables

The most widely used insulation compounds are based on PE, PP, silicone rubber, EPR, PVC, and fluoroplastics. Polyethylene (thermoplastic or cross-linkable) is used because it is lightweight, water-resistant, and easy to strip, and has low dielectric constant and power loss. It is used, especially in foamed form, to make computer and TV coaxial cables. Polypropylene has very good abrasion resistance and its heat resistance is better than that of polyethylene. It has low electrical losses but since it is relatively stiff its use is rather limited. Silicone rubber-based compounds are used to produce wires used at high temperatures (due to its good aging properties) and for special fire-resistant application, due to its char formation during fire, such as for Navy shipboard wires. Ethylene–propylene rubber-based compounds have some use in low voltage cables but are much more popular in manufacturing power cables.

Examples of fluoroplastics include polytetrafluoroethylene (PTFE), fluorinated ethylene propylene (FEP), ethylene–chlorotrifluoroethylene (ECTFE), ethylene–tetrafluoroethylene (ETFE), poly(vinylidene fluoride) (PVDF), etc (see FLUORINE COMPOUNDS, ORGANIC). These polymers have outstanding electrical properties, such as low power loss and dielectric constant, coupled with very good flame resistance and low smoke emission during fire. Therefore, in spite of their relatively high price, they are used extensively in telecommunication wires, especially for production of plenum cables. Plenum areas provide a convenient, economical way to run electrical wires and cables and to interconnect them throughout nonresidential buildings (14). Development of special flame-retardant low smoke compounds, some based on PVC, have provided lower cost competition to the fluoroplastics for indoors application such as plenum cable, Riser Cables, etc.

Poly(vinyl chloride). PVC has intrinsic resistance to fires, oils, most chemicals, ozone, and sunlight. Due to its natural stiffness and rigidity, it cannot be used as is but is compounded with various ingredients, especially plasticizers, in order to obtain flexibility as well as other properties. In compounded form, PVC is used either as insulation in areas where its relatively high dielectric constant and dielectric power loss is acceptable (such as wires used for audiotransmission, low voltage building, and portable), or as jacketing for a large variety of cables including power cables. UL specifications define certain temperature rating criteria based on physical aging characteristics of PVC compounds (15). Other UL test criteria for PVC compounds used in wires and cables includes the cold-bend test, deformation test, heat–shock test, vertical flame test, horizontal flame test, tray–cables flame test, smoke emission test, dielectric voltage–withstand, etc.

In a flexible PVC compound, ingredients in the recipe are chosen based on cost and or their contribution to physical and other properties and performance. Typical ingredients (16,17) are stabilizers, fillers, plasticizers, colorants, and lubricants.

Plasticizers. Monomeric (mol wt 250–450) plasticizers (qv) are predominantly phthalate, adipate, sebacate, phosphate, or trimellitate esters. Organic phthalate esters like dioctyl phthalate (DOP) are by far the most common plasticizers in flexible PVC. Phthalates are good general-purpose plasticizers which impart good physical and low temperature properties but lack permanence in hot or extractive service conditions and are therefore sometimes called migratory plasticizers. Polymeric plasticizers (mol wt up to 5000 or more) offer an improvement in nonmigratory permanence at a sacrifice in cost, low temperature properties, and processibility; examples are ethylene vinyl acetate or nitrile polymers.

Stabilizers. Heat stabilizers (qv) are included in PVC compounds to counteract the internal generation of hydrogen chloride as well as the external degradative effect of heat. Due to environmental considerations, there is a trend toward decreasing and even avoiding the use of stabilizers based on heavy metals, eg, lead.

Colorants. Pigments are the main colorants used in PVC, but some dyes are also employed (see COLORANTS FOR PLASTICS).

Lubricants. Process aids or lubricants promote smooth and rapid extrusion and calendering, prevent sticking to extruders or calender rolls, and impart good release properties to molding compounds. In some cases use of lubricants allow slightly lower processing temperatures (see VINYL POLYMERS).

Fillers. These are used to reduce cost in flexible PVC compounds. It is also possible to improve specific properties such as insulation resistance, yellowing in sunlight, scuff resistance, and heat deformation with the use of fillers (qv). Typical filler types used in PVC are calcium carbonate, clays, silica, titanium dioxide, and carbon black.

The PVC formulations shown in Table 2 represent typical compounds used by the wire and cable industry. PVC compounders have developed new PVC-based formulations with very good fire and smoke properties (can pass the UL 910 Steiner Tunnel test) that compete with the more expensive fluoropolymers. These can be used in fabricating telecommunication cables usable for plenum area applications.

Table 2. Wire and Cable Insulation PVC Formulations, Parts by Weight

Component	Low cost for low temperature	For high temperature applications	Nonmigratory plasticizer
-----------	------------------------------	-----------------------------------	--------------------------

PVC	100	100	65
dioctyl phthalate	70		
diundecyl phthalate		25	
trimellitate		25	
nitrile rubber			35
lead-based stabilizer	7	6	5
clay	20	20	
calcium carbonate	10	10	
antioxidant system	0.25	0.5	2
antimony trioxide			3
stearic acid			0.25
acrylic-based modifier			5

Magnet Wires. Magnet wires can be classified as coated wires, coated wires with fibrous wrappings, and wires with impregnated fibrous wrappings; the last two categories are older technologies. Wires coated with only an organic coating are frequently referred to as enamel wires or simply coated wires. The organic coating (one or multilayers) is applied directly to the conductor and is a dielectric material.

Examples of thermoplastic coatings are fluoropolymers, eg, Teflon or polyamides, eg, nylon. Thermosetting coatings are more resistant to cut-through and have superior resistance to heat and solvents. The silicones, polyimides, and fluorocarbons are best suited for very high temperatures applications, the polyurethanes for ease of removal, and epoxies for solvent and chemical resistance. Several other polymers are also used to coat the magnet wires. A summary of their advantages and limitations are given in Table 3 (18).

Power Cables. The materials mostly used to produce power cables are ethylene copolymers loaded with conductive carbon black for semiconductive shielding layers, polyethylene or ethylene–propylene rubber-based compounds as insulations, and either thermoplastic materials (eg, polyethylene, PVC) or thermosetting (based on chlorinated polyethylene (CPE), chlorosulfonated polyethylene (CSPE), chloroprene, etc) for jackets.

Table 3. Properties of Coated Wires^a

Coating	Thermal rating, °C	Advantages	Limitations
poly(vinyl formal)	105	toughness, dielectric strength; compatible with other coatings; heat-shock resistant	crazes in polar solvents
polyurethane	105	dielectric strength, chemical resistance, moisture, and corona resistance; compatible with solvents and chemicals; solderable without stripping	low thermal resistance
polyamide (nylon)	105	toughness, dielectric strength, solvent resistance; solderable; good windability	high moisture absorption; high electrical loss at all frequencies
poly(vinyl formal)–poly(vinyl butyral)	105	bondability, dielectric strength; heat–shock resistant	vibration; high mechanical stress
polyester	155	toughness, dielectric strength, chemical resistance, cut-through resistance	hydrolyzes in moist sealed atmosphere
polytetrafluoroethylene (Teflon)	200	thermal stability, chemical stability, dielectric strength; low dielectric constant	high abrasion; high gas permeability; cold flow; poor adhesion
polyimide	220	high overload resistance, thermal resistance, chemical stability, radiation resistance; high cut-through resistance	stripping difficulty; crazes in some solvents

^a Ref. 18.

Insulation. Cross-linked polyethylene (XPLE) and ethylene–propylene rubber (EPR), both thermosets, are the primary extruded dielectrics used in medium and high voltage power cables (Table 4).

Table 4. Components Used in Power Cable Insulations Based on EPR, Parts by Weight

Component	Low (to 5 kV)	Medium (to 35 kV)	High (to 138 kV)
EPR	100	100	100
low density PE	0–20	0–20	0–20
paraffinic wax	0–5	0–5	0–5
stearic acid	1–3		
calcium carbonate	50–100		
calcined clay	100–200		
silane treated clay		100–150	50–100
paraffinic oil	50–150	0–30	0–20
zinc oxide	0–5	0–5	0–5
lead oxide		0–5	0–5
antioxidants	1–2	1–2	1–2
coupling agent	1–2		
peroxide	3–6	3–6	3–6

High dielectric strength and very low electric conductivity make polyethylene an outstanding insulator for electric power cable at low voltages as well as high voltages used by transmission cables. Polyethylene is also the most suitable dielectric for all types of high frequency cables because of its low dielectric loss at high frequencies and its remarkable mechanical properties.

The power factor of polyethylene which provides the measure of the power loss in the insulated conductor increases slightly with an increase in the temperature of the atmosphere or the electrical equipment, both of which may fluctuate widely. It also increases slightly with an increase in the humidity of the surroundings.

Improved heat resistance is the most important advantage of cross-linked polyethylene (XLPE) over thermoplastic polyethylene. A power cable

with XLPE insulation can operate at conductor temperatures of 90°C. Since conductor temperature is proportional to the amount of current sent through the cable, more power can be sent through an XLPE cable than through a noncross-linked cable of the same size. Thus in heavily populated areas, fewer or smaller XLPE cables can be installed. In appliance wire applications, cross-linking allows compounds to be formulated for 125°C service temperatures, well above the melting point of the noncross-linked base resin.

Compared to a typical cross-linked polyethylene-based compound, the typical EPR-based compounds used for medium voltage cables contain much larger amounts of ingredients. Besides being higher in cost, when compared to XLPE the EPR-based insulations display certain inferior electrical properties, such as higher dielectric loss, lower dielectric and impulse strength, especially when measurements are done on newly produced cable, before field operation. However, the longtime field service records have shown numerous positive features for the power cables insulated with compounds based on EPR that are making them attractive to customers interested in cables with a long life history in the field (19).

Compared to XLPE, the EPR-based insulation compounds used in power cables have the following characteristics: greater flexibility and ease of installation; easier splicing and terminating in all weather; lower coefficient of thermal expansion at high temperatures, generated during emergency overloads and short circuits, thus lower tendency to separation between insulation and insulation shield layers as well as between the components of the cable and of the premolded splicing kits that typically are based on EPR; superior resistance to degradation caused by partial discharges in voids within the insulation or at the interface between the insulation and the shielding layers of the power cables; and less tendency to water treeing degradation and failure, possibly due to the lower crystallinity of EPR and to higher filler content vs polyethylene-based formulations of EPR compounds (19).

Besides using polyethylene and EPR as materials of choice for the insulations of power cable, there are a few other technologies that are less popular but still in use. In pressurized filled cables, the cable is kept full of oil under pressure by oil reservoirs connected to cables. Solid paper insulated cables, where the oil is impregnated into the paper tape during manufacturing, are used for low voltages due to corona effects that may occur at high voltages in the voids that may exist in layers. However, for high voltages (up to 230 kV) the oil is kept under pressure to fill the eventual voids.

Extruded materials are used for power cables from 5 kV up to 138 kV for underground distribution and transmission lines; the 230 kV cable is still in infancy. Compared to the low voltage cables (up to 5 kV) that use simpler materials, the medium voltage cables (5 to 35 kV), the high voltage power cables (up to 138 kV), and the very high voltage cables (230 kV and higher) contain specially developed materials due to the more difficult and special applications concerns.

Shields. Power cable (conductor) shields provide a smooth, continuous, conductive, and isopotential interface between the conductor and insulation (Fig. 5). The geometry of the conductor strands permits air gaps between the outer wires of the stranded conductor and the inner surface of the extruded insulation. Without a stress control layer, excessive electric gradients can cause partial discharges within these gaps that harm the insulation. There are two design approaches: most shields are either semiconductive shields that use large amounts of carbon black mixed in polymeric-based formulations, or stress-relieving shields that are based on materials with high dielectric constant. Brand names for the latter are PermaShield or Emission Shield.

The interface between conductor shield and insulation is the region of the highest stress in the cable insulation structure. Any imperfections at this interface, especially sharp protrusions of the conductor shield into the insulation, will cause high local electrical stress that may reduce the dielectric strength of finished cable. Calculation of the stress enhancement, for a 15 kV cable with a 4.4 mm (175 mil) insulation thickness, indicates that the common round 50 μm (2 mil) radius protrusions increase the electrical stress by a factor of 30 and a sharp 5 μm protrusion will increase the electric stress by as much as 210 times (11,20).

Trees originating at a shield–insulation interface are mostly due to the existence of protrusion from the shields. They are referred to as vented trees; if moisture is present, they are called vented water trees. Particulate contaminants present in the insulation, and waterborne ionizable materials that find their way into the insulation, are also causes of tree formation.

The carbon black in semiconductive shields is composed of complex aggregates (clusters) that are grape-like structures of very small primary particles in the 10 to 70 nanometer size range (see CARBON, CARBON BLACK). The optimum concentration of carbon black is a compromise between conductivity and processibility and can vary from about 30 to 60 parts per hundred of polymer (phr) depending on the black. If the black concentration is higher than 60 phr for most blacks, the compound is no longer easily extruded into a thin continuous layer on the cable and its physical properties are sacrificed. Ionic contaminants in carbon black may produce tree channels in the insulation close to the conductor shield.

The conductive carbon black particles suspended in the compound's polymeric base may assume configurations that will create high stress points at the interface between the conductor shield and the insulation. These points, similar to protrusions, can be very sharp and cause localized voltage stresses which significantly exceed the electric stresses calculated for a uniform surface. These extremely high local voltage stresses, caused by protrusions and or carbon black particles suspended in the semiconducting compound, can initiate cold electron emission from carbon black particles and or initiate partial discharges, which in turn may cause insulation breakdown (20).

Insulation Shields. The insulation shield is a layer applied over the insulation (see Fig. 5). It plays much the same role as the conductor shield in protecting the insulation from the damaging effects of ionization at the outside of the insulation surface, therefore it too must always remain in intimate contact with the insulation and be free of voids and defects at the interface. As an integral component of cable grounding, the insulation shield must be a resistive shield, providing a uniform ground around the insulation during field service; it also contributes to the grounding of the cable during switching surges, short circuits, or lightning strikes.

The electric stress at the interface between the insulation and the insulation shield is less than at the conductor shield–insulation interface.

Most medium voltage cables are made with insulation shield layers that are bonded but easily stripped from the insulation in order to avoid pockets of air at the interface and at the same time to allow easy field handling for termination and splicing (during installation).

The cables designed for use at voltages over 49 kV require that the conductor and insulation shields be firmly bonded to the insulation in order to avoid any possibility of generating corona at interfaces; strippable insulation shields are not accepted. The AEIC specifications for cables rated for 59–138 kV require a volume resistivity of one order of magnitude lower than for the medium voltage cables.

The most important parameter that affects the resistivity is the amount of carbon black particles, and of secondary importance is the type and especially the shape of the carbon black particles. The susceptibility of the carbon black to oxidation may possibly lead to high resistivity of insulation shields. The type of polymer used in a semiconducting material is also an important parameter that can affect resistivity.

Processing conditions also significantly affect the lengths and numbers of continuous carbon black chains, therefore the semiconducting shields must be applied with a minimum of residual mechanical stress.

Jacketing Materials. Besides the metallic protective coverings (based on aluminum, copper and copper alloys, lead, steel, and zinc), the most popular jacketing materials are based on polymeric materials that can be either thermoplastic (with limited high temperature use) or thermosetting.

Polyethylene has been the most popular material for power cable jacketing due to its moisture resistance, abrasion resistance, toughness, and especially its relatively low cost. The original low density polyethylene (LDPE) has been replaced by high density polyethylene (HDPE), and by the newer linear low density polyethylene (LLDPE). The main reasons for this change are its superior flexibility and environmental stress–crack resistance as well as lower shrinkage tendency when compared to HDPE. Poly(vinyl chloride) (PVC) is still widely used where the flame retardancy and chemical resistance is important. Polyamides are limited to smaller size wires when the mechanical toughness is required. Polyurethanes are used for areas where the abrasion resistance is important. Thermoplastic elastomers (TPE) that are typically blends of thermoplastic polymers such as polypropylene with elastomers such as EPR, confer a combination of physical strength and flexibility (see ELASTOMERS, THERMOPLASTIC).

Ethylene vinyl acetate (EVA) polymers are used in thermoplastic and thermosetting jacketing compounds for applications that require flame retardancy combined with low smoke emission during the fire as well as the absence of halogen in the composition.

Thermosetting jackets are still used in applications that require high temperature rating. Polychloroprene, eg, Neoprene, was the first synthetic rubber used in wire and cable jackets due to its good resistance to sunlight, fire, and chemicals. Chlorosulfonated polyethylene (CSPE), eg, Hypalon, has replaced most Neoprene due to its superior heat, light, and moisture resistance combined with easier processibility. Chlorinated polyethylene, eg, Tylin, is very similar to CSPE polymer except it does not contain inherent sulfur; therefore, the vulcanized CPE-based compounds have good colorability and also can be used in contact with bare copper cables such as in manufacturing heater cords. Jacket materials that contain nitrile rubber are used in compounded

form with various ingredients in jacketing applications that require oil resistance or resistance to color fading of colored jackets. Sometimes the nitrile rubber-based elastomeric compounds contain PVC as a component in order to improve their ozone resistance (see ELASTOMERS, SYNTHETIC NITRILE RUBBER).

BIBLIOGRAPHY

"Dielectrics" in *ECT* 1st ed., Vol. 5, pp. 51–75, by E. B. Baker and W. C. Goggin, The Dow Chemical Co.; "Insulation, Electric" in *ECT* 2nd ed., Vol. 11, pp. 774–801, by J. Swiss, Westinghouse Electric Corp.; "Properties and materials" under "Insulation, Electric" in *ECT* 3rd ed., Vol. 13, pp. 534–563, by R. N. Sampson, Westinghouse Electric Corp.; "Cable Coverings" in *ECT* 1st ed., Vol. 2, pp. 697–716, by E. D. Youmans, The Okonite Co.; "Electric Wire and Cable Coverings" under "Insulation, Electric" in *ECT* 2nd ed., Vol. 11, pp. 802–822, by J. Hogan, The Okonite Co.; "Wire and Cable Coverings" under "Insulation, Electric" in *ECT* 3rd ed., Vol. 13, pp. 564–590, by J. Hogan, The Okonite Co.

1. *Wire Industry News*, 9 (Oct. 24, 1991).
2. U.S. Department of Commerce, Bureau of Census, *Insulated Wire and Cable*, MA33L, Washington, D.C., 1990.
3. Du Pont Co. guide, *How to Specify, Bid and Install Plenum Cable*, Wilmington, Del., 1986, pp. 6–16.
4. G. Baker, *LAN Magazine*, 38–42 (1986).
5. *Belden Company Catalog for Electronic Wire and Cable*, Richmond, Ind., 1989.
6. Kerite Co., *Kerite Cable-Dependable Power from Source to Load*, Seymour, Conn.
7. Independent Cable Co. Inc., *Cable Guide Catalog*, Hudson, Mass.
8. Olflex Co., *Wire and Cable Catalog U29*, Fairfield, N.J., 1985.
9. Kerite Co., *Distribution Cable and Power Cables*, data catalogues, Seymour, Conn.
10. R. T. Vanderbilt Co., *Rubber Handbook*, Norwalk, Conn., 1991, pp. 701–715.
11. Union Carbide Corp., *Kabelitems Wire and Cables No. 150, 152, Treeing Update*, Danbury, Conn.
12. Union Carbide Corp., *Kabelitems Wire and Cables No. 157, A Critical Comparison of XLPE and EPR for use as Electrical Insulation on Underground Power Cables*, Danbury, Conn.
13. Union Carbide Corp., *Kabelitems Wire and Cables No. 160, Long-life Insulation Compounds Update*, Danbury, Conn.
14. AlliedSignal Inc., *Technical Bulletin, Engineering Plastics, HALAR ECTFE Fluoropolymer Resins Comparison Data*, Morristown, N.J.
15. Underwriters Laboratories, *Spec. Subject 13 for Power-limited Circuit Cable Class 2*, (NEC article 725), Melville, N.Y.
16. Goodyear Co., *Chemigum Powder Nitrile/PVC Alloys*, compounding guide, Akron, Ohio, 1988.
17. B. F. Goodrich Co., *Typical Properties of Geon Vinyl Compounds for Wire and Cables Insulation and Jacketing*, bulletin, Cleveland, Ohio, 1991.
18. J. J. Licari, *Handbook of Polymer Coatings for Electronics*, Noyes Publications, Park Ridge, N.J., 1990, pp. 250–277.
19. M. Brown, *IEEE Electrical Insulation Magazine*, 21–26 (1991).
20. N. M. Burns, R. M. Eichorn, and C. G. Reid, *IEEE Electrical Insulation Magazine*, 8–24 (1992).

General References

J. Irvine and S. Hookanson, "Development of a Flame-Retardant Premium Cable," *Wire J. Int.*, 76–79 (1983).

B. J. McPartland, "The Materials and Makeup of Type MV Cable," *EC&M*, 71–75 (1987).

R. B. Blodgett, "Ethylene–Propylene Rubber and Crosslinked Polyethylene as Insulations for 90°C Rated Medium Voltage Cables," *Rubber Chem. Technol.*, **52**, 410–424 (1978).

A. Barlow, "The Chemistry of Polyethylene Insulation," *IEEE Electrical Insulation Magazine*, 8–19 (1991).

R. J. Arhart, "The Chemistry of Ethylene–Propylene Insulation," *IEEE Electrical Insulation Magazine* (1993).

R. Bartnikas, "Engineering Dielectrics Vol. II A—Electrical Properties of Solid Insulating Materials: Molecular Structure and Electrical Behavior," *ASTM Special Technical Publication 783*, 1983, Chaps. 1–5, pp. 3–515.

G. Elber, "Vinyl," *Plastics Design Forum*, 48–54 (1993).

"Non-Lead PVC Stabilizers Closing Performance Gap in Wire," *Plastics Technology Magazine*, 35–36 (1993).

Armand Moscovici
The Kerite Company

INSULATION, THERMAL

Insulation

A variety of cellular plastics exists for use as thermal insulation as basic materials and products, or as thermal insulation systems in combination with other materials (see FOAMED PLASTICS). Polystyrenes, polyisocyanurates (which include polyurethanes), and phenolics are most commonly available for general use, however, there is increasing use of other types including polyethylenes, polyimides, melamines, and poly(vinyl chlorides) for specific applications.

Originally polystyrenes and polyurethanes were developed for applications involving severe environmental conditions. In the 1990s, however, primary applications include use in refrigerators and freezers, where cellular plastics account for over 90% of total insulation; in the building envelope, ie, from foundation to roof, 70%; and in walls, sheathings, and basements, ca 40%. Other uses include pipelines, refrigerated transportation, chemical processing, road and runway beds, and cryogenic applications.

Foamed plastics (qv) were developed in Europe and the United States in the mid-to-late 1930s. In the mid-1940s, extruded foamed polystyrene (XEPS) was produced commercially, followed by polyurethanes and expanded (molded) polystyrene (EPS) which were manufactured from beads (1,2). In response to the requirement for more fire-resistant cellular plastics, polyisocyanurate foams and modified urethanes containing additives were developed in the late 1960s; urea-formaldehyde, phenolic, and other foams were also used in Europe at this time.

The newer open-cell foams, based on polyimides (qv), polybenzimidazoles, polypyrones, polyureas, polyphenylquinoxalines, and phenolic resins (qv), produce less smoke, are more fire resistant and can be used at higher temperatures. These materials are more expensive and used only for special applications including aircraft and marine vessels. Rigid poly(vinyl chloride) (PVC) foams are available in small quantities mainly for use in composite panels and piping applications (see FLAME RETARDANTS; HEAT RESISTANT POLYMERS).

Cellular plastics have been used extensively for low temperature applications (3–5); however, uses for cellular plastic insulations, particularly molded and extruded polystyrenes, polyisocyanurates, and phenolics, have expanded into the building arena (6–10) as a result of the energy crisis of the early 1970s, when the use of increased insulation for energy conservation became more economically attractive. All national codes and standards specify the need for levels of insulation in the building envelope (foundation to roof) dependent on the climatic region, and building applications in the 1990s account for over 80% of the total volume of cellular plastic materials used for insulation purposes (11) (see BUILDING MATERIALS).

Function of Thermal Insulation

Three basic mechanisms of heat transmission occur in thermal insulation: radiation (electromagnetic waves) (12), conduction (atomic or molecular collisions), and convection (fluid motion). Radiation is the primary mode as shown in Figure 1. The function of thermal insulation is to minimize and control these modes. This is accomplished primarily by introducing low emittance–high reflectance barriers to attenuate radiation; incorporating a large number of small, low density, low thermal conductivity elements to minimize solid conduction and convection; and including high density, low thermal conductivity gas, or evacuation of encapsulated systems to minimize convection and gas conduction.

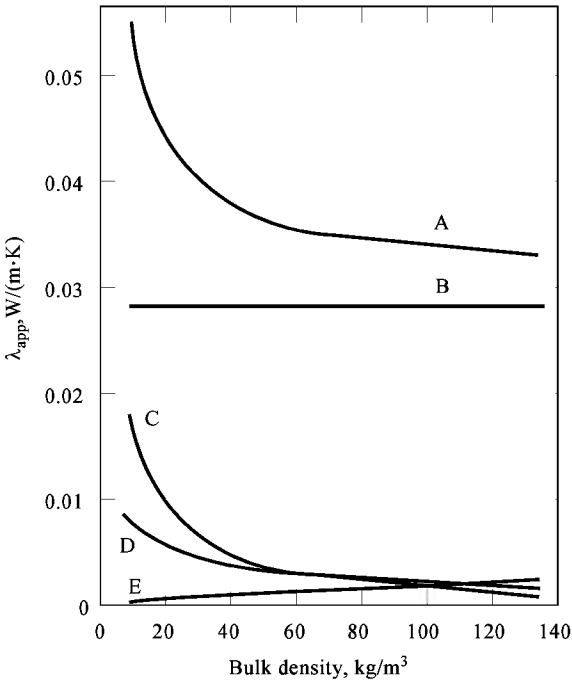


Fig. 1. Thermal conductivity components vs density for a typical thermal insulation material at 300 K: A, total conductivity; B, air conduction; C, radiation; D, convection; and E, conduction of solid polymer.

Although thermal performance is a principal property of thermal insulation (13–15), suitability for temperature and environmental conditions; compressive, flexure, shear, and tensile strengths; resistance to moisture absorption; dimensional stability; shock and vibration resistance; chemical, environmental, and erosion resistance; space limitations; fire resistance; health effects; availability and ease of application; and economics are also considerations.

Cellular Plastics as Insulation

A low (<0.4 W (m·K)) thermal conductivity polymer, fabricated into a low density foam consisting of a multitude of tiny closed cells, provides good thermal performance. Cellular plastic thermal insulation can be used in the 4–350 K temperature range. Cellular plastic materials have been developed in

various types, eg, open-cell, closed-cell (the most energy efficient), and closed-cell containing gases with a thermal conductivity approximately one-half that of air.

Organic foams have excellent thermal insulation characteristics including high strength-to-weight ratio, versatility, and cost effectiveness. They are also self-supporting and can support some load, depending on material and density. Organic foams are available in rigid, semirigid, flexible, and reinforced forms, can be fabricated as board and pipe stock by extrusion, expansion, and molding, and can be applied by spraying, foaming, or frothing. Pellets and beads are used as loose fill.

The drawbacks of cellular materials include limited temperature of applications, poor flammability characteristics without the addition of fire retardants, possible health hazards, uncertain dimensional stability, thermal aging and degradation, friability, and embrittlement due to the effects of uv light (3,6,15).

Materials

Polystyrene. Polystyrene exists in two forms, namely extruded (XEPS) and expanded or molded (EPS). Extruded polystyrene is manufactured by passing a hot mixture of polystyrene, solvent, and pressurized gas, serving as a blowing agent, through an orifice. The gas expands resulting in a fine closed-cell (>90%) structure. The EPS form is fabricated by heating performed polystyrene beads and a blowing agent, such as pentane, in a mold. The vapor pressure of the gas causes the beads to expand thus producing a predominantly closed-cell material.

Polystyrene foams are used in both residential and commercial industrial buildings for insulating all parts of the building envelope (see STYRENE PLASTICS). Where it is used as exterior sheathing and backing of siding, polystyrene foam is often faced with a reflective foil and used in conjunction with an airspace to enhance the total thermal resistance of the system. The high resistance to water absorption of XEPS, as well as its improved mechanical strengths due to a higher density range of use, make it suitable for applications which include building perimeters, foundations, and upside-down roofs generally referred to as protected membrane roof (PMR) (see ROOFING MATERIALS). The expanded beads alone are used as a loose fill insulation particularly for cavities in masonry constructions.

Polyisocyanurate Including Polyurethane. Polyurethane foams are formed by the reaction of isocyanates and polyfunctional alcohols or polyols in the presence of a suitable blowing agent. Polyisocyanurates are manufactured from isocyanates, a catalyst, and similar blowing agents. A >90% closed-cell rigid foam is formed by choosing the appropriate isocyanate functionality, alcohol, and molecular weight (see ISOCYANATES, ORGANIC; URETHANE POLYMERS).

The foams are available in several forms. Slab stock is manufactured by mixing components and continuously feeding metered mixture onto a conveyor. For laminated material, mixture is fed between impermeable or low permeable facings such as aluminum foils or reinforced papers or plastics. The double-bond lamination process is used extensively with flexible laminates such as paper and reinforced polymer film. The materials can also be formed *in situ* by manual or automatic dispensing of the material components into a closed cavity, eg, appliance components and preformed building panels (16). Finally, a significant and increasing amount of foam is fabricated in the field by directly spraying onto any appropriate clean dry surface, especially roof decks (17).

These foams are used for board stock in commercial and industrial buildings as insulation for internal cavity and external walls, roof, floor, and foundations. Spray-applied foam, covered subsequently with one of a variety of protective coatings, is widely used for large roofing applications and has limited use as external walls. For residential buildings, the principal use is as external sheathing board.

Phenolic. The reaction product of a phenol and an aldehyde with a blowing agent in the presence of a catalyst is a phenolic foam. The manufacturing process for faced and unfaced materials is somewhat similar to that for urethane materials and results in a product having greater than 90% closed-cell content. Some open-cell or partially closed-cell materials are manufactured when a hydrocarbon blowing agent is used. Phenolic foams are used mainly for roofing insulation; application is limited as sheathing products for external wall insulation for building applications, and for shaped parts such as pipe and block insulation for industrial applications.

Polyethylene. This is essentially a closed-cell insulation manufactured at 448 ± 2 K by an extrusion process. A blowing agent and nucleating agent are employed to control the cell size, and primary use is in insulating pipelines for hot and chilled water lines, air conditioning, and processing systems.

Polyimide and Melamine. These are both low density, essentially open-cell foams used as pipe insulations, particularly those involved with fluids operating at temperatures up to 530 K. Because these foams operate at higher temperatures and have improved flammability characteristics compared to other foams, they are also used for some aircraft and marine applications.

Other.

Urea–Formaldehyde and Urea-Based. In the 1970s and early 1980s, urea materials were in general use particularly for direct field retrofitting of cavity wall construction of wood frame and masonry. However, because of formaldehyde odor and excess shrinkage under specific conditions, this cellular plastic has limited use as an insulation.

Properties

Significant properties of insulation (Table 1) include thermal conductivity, fire resistance, and minimal production of toxic gases primarily during combustion. Other criteria include water-vapor permeability, resistance to water absorption, and dimensional stability over prolonged periods of submission to extreme environments.

Table 1. Typical Properties of Cellular Plastic Materials Used as Thermal Insulation

Property	ASTM method	Polyisocyanurate	XEPS ^a	EPS ^b	Polyimide	Polyethylene	Phenolic
density, kg m ⁻³	C591	30–40	30–48	12–30	8–12	21–32	45–60
closed-cell content, %		>90	>90	>90	<10	>90	>90
water-vapor permeability ^c	C355	2–3	0.4–0.15	1–4	high	0.02	<1
water absorp-tion, vol %	C272	2–5	0.15	2–4		<1	<2
thermal expansion × 10 ⁻⁶ °C ⁻¹	E228	30–40	30–40	30–45	30–40	30–50	20–40
heat capacity, J (kg·K) ^d	C351	1500	1200–1300	1200–1300			2000
thermal con-ductivity, W (m·K)	C518	0.026 ^e 0.020 ^f	0.028	0.038–0.033	0.043	0.035	0.018
fire resistance	E136	combustible	combustible	combustible	combustible	combustible	combustible
flame spread ^c	E84	25–50	5–15	10–25	12	<25	20–25
smoke	E84	155–50055–200	10–40	125	7	<50	5–15

development ^a toxicity		toxic gases when burned	CO when burned <2	CO when burned <1	CO when burned	CO whenburned	CO when burned <1
dimensional stability, vol %	D2126	0–12					
upper tempera-ture limit, °C		120	75	75	260	180	150

^a Extruded.

^b Molded.

^c Arbitrary (qualitative) units given in ASTM test.

^d To convert J to cal, divide by 4.184.

^e Aged unfaced.

^f Impermeable skins.

Thermal Conductivity and Aging. Thermal performance is governed by gas conduction and radiation (18–20). In most cellular plastic insulations, radiation is reduced because normal densities of use are 4–50 kg m^{−3} and the average cell size is <0.5 mm. For open-cell and other materials containing air (at 24°C, $\gamma = 0.025$ W (mK)) this results in total values of λ at 0.029–0.0039 W (mK).

However, for closed-cell extruded polystyrene, polyisocyanurate, and phenolic foams containing high molecular weight and other low thermal conductivity gaseous blowing agents (at 24°C, $\gamma = 0.025$ W (mK)), the initial values of λ , as blown, are between 0.013 and 0.02 W (mK). These values can be maintained only if the aging process cannot occur, ie, air cannot diffuse into the cells or the blowing agent cannot diffuse out or partially dissolve into the polymer matrix. To accomplish this, the materials must be contained with impermeable, thick membranes such as metal sheets or hole-free foils well adhered to the cellular polymer. Other more permeable facings, especially if not adhered to the foams faces, allow the aging process to take place.

In general, unless carbon dioxide or another low molecular weight gas such as pentane is used as the blowing agent, the air components diffuse inward much faster than the outward diffusion of the blowing agents. The overall process is represented in Figure 2 which is based on measurements on thin and thick specimens. It occurs in two thickness-dependent stages, a primary (<180–360 d) and secondary (>10–20 yr) phase, generally with a clear transition point. The process is complex and each stage occurs at a rate dependent on the polymer type, the structure of the foam, the temperature, the gas type, and its concentration and pressure. The permeation rate, P , requires knowledge of the diffusion coefficient, D , and the gas solubility, S , of the polymer (20).

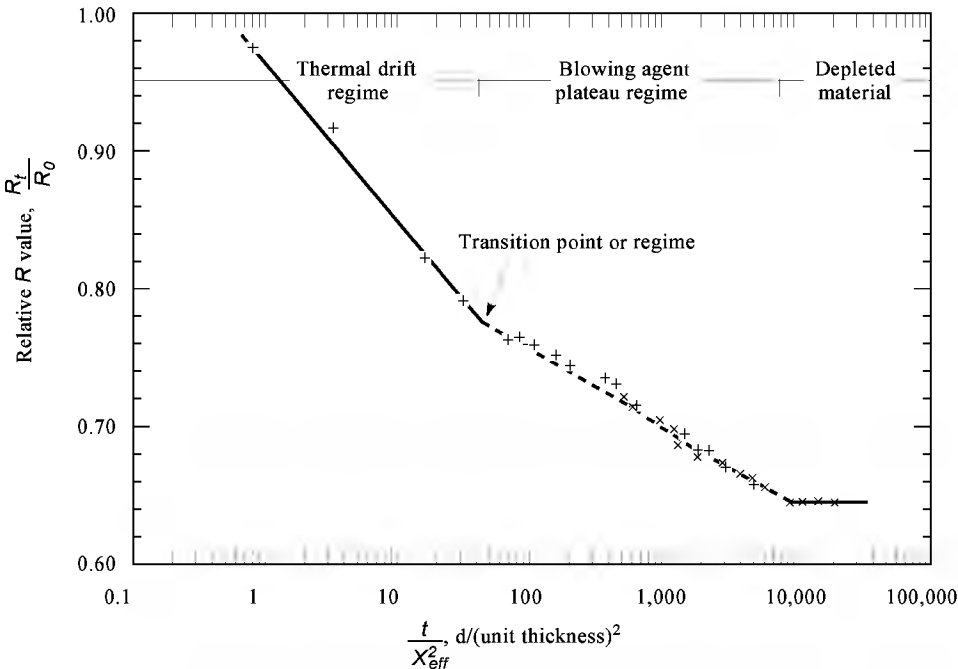


Fig. 2. Change in insulation value with time for insulation of various thicknesses. Slice thickness = 2.5–6 mm; R_t and R_0 are defined as thermal resistance at time t and zero, respectively.

Some practical results for different types and forms of cellular plastic that do exhibit aging are shown in Figure 3. It is necessary to use such aged values for true performance characteristics especially for specifications and energy use purposes. The more recently developed phenolic foams do not have the same aging characteristics as the extruded polystyrene and polyisocyanurates, probably due to higher density, smaller cell size, thicker cell walls, and lower gas permeability and solubility in the polymer. Similarly the somewhat improved performance of spray-applied urethane is due to a more uniform cell size and structure combined with the separately applied protective membrane on the outer surface (17). In addition, formation of intermediate higher density skin layers takes place during the application process which normally involves several separate passes of 10–12 mm thick sections to form the required thickness.

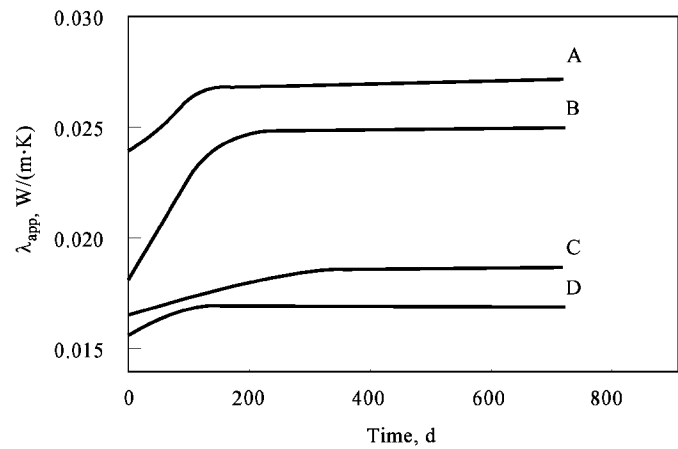


Fig. 3. Aging effect on thermal conductivity of cellular plastics: A, extruded polystyrene; B, unfaced polyurethane; C, unfaced phenolic; and D, polyurethane with thick steel skins.

Many cellular plastic products are available with different types of protective faces, including composite metal and plastic foils, fiber-reinforced plastic skins, and other coatings. These reduce but do not eliminate the rate of aging. For optimum performance, such membranes must be totally adhered to the foam, and other imperfections such as wrinkles, cuts, holes, and unprotected edges should be avoided because they all contribute to accelerated aging.

Blowing Agents and Accelerated Aging Testing. Until the late 1980s, the fully halogenated chlorofluorocarbons (CFCs), primarily CFC11 and CFC12, were the predominant blowing agents used to produce closed-cell cellular plastic insulations. However, during the 1980s, there had been growing evidence that these gases were contributing both to a depletion of the ozone layer and to an increase in the so-called greenhouse warming effect (21). As a result, the Montreal Protocol, an international agreement whereby CFCs were to be phased out gradually, was developed in 1987 (22). It has been subsequently revised to ensure CFC elimination by the year 2000 at the latest. In addition, there was a nonbinding declaration of intent that the hydrochlorofluorocarbons (HCFCs), the most widely touted substitutes for CFCs, also be phased out no later than 2020 (see FLUORINE COMPOUNDS, ORGANIC FLUORINATED ALIPHATIC). Attention must also be paid to minimizing the release of either gas when outdated material is disposed of or recycled (see RECYCLING, PLASTICS).

Research and development programs have been initiated by the cellular plastics industry to develop viable substitute blowing agents. These must have similar or improved properties to their CFC counterparts at a reasonable cost. Emphasis was initially placed on HCFC 123 and HCFC 141b, both having much shorter lifetimes and considerably less effect (up to 50 times) on ozone layer depletion (22). However, various options, including gas mixtures, water, or CO₂ blown foams, continue to be studied ultimately to eliminate all CFCs and HCFCs.

The search for alternative blowing agents has necessitated a significant change in testing requirements to provide some assessment of aged value. Accepted accelerated testing procedures normally use >25 mm thick specimens and time exposure periods of 180 days at 24°C or 90 days at 60°C to indicate some amount of aging. Measurements of thermal resistance are made after such conditioning to give curves such as those shown in Figure 3. However, there is both the need to obtain more realistic longer term (>20 yr) values and to test much more rapidly, since the development of safe blowing agent substitutes cannot wait on determining realistic aged values. Diffusion coefficients are not a strong function of temperature nor are all gases changed equally, and elevated temperature exposure can damage foam structure.

An alternative method known as slicing and scaling has been developed (23,24). In this, the rate of diffusion is determined on a thin specimen (6–10 mm thick) and a scaling factor *S* used to relate the results to a thick specimen. For a material satisfying the requirements of a constant diffusion and constant initial pressure, *p*, the same ratio of time:thickness² provides the same values of *p* and *λ*. Thus the thermal resistance of a specimen of thickness *l*₁ at time *t*₁ can be obtained by conditioning a specimen of thickness *l*₂ over a time *t*₂ given by

$$t_2 = t_1 \left[\frac{(l_2)^2}{(l_1)^2} \right] \text{ ie, } t_2 = t_1 S$$

This technique reduces testing times significantly and provides reliable results for >20 years material. The values plotted in Figure 2 (25) are an illustration of the viability of this technique as a means to provide realistic long-term thermal performance values (21).

Thickness. The traditional definition of thermal conductivity as an intrinsic property of a material where conduction is the only mode of heat transmission is not applicable to low density materials. Although radiation between parallel surfaces is independent of distance, the measurement of *λ* where radiation is significant requires the introduction of an additional variable, thickness. The thickness effect is observed in materials of low density at ambient temperatures and in materials of higher density at elevated temperatures. It depends on the radiation permeance of the materials, which in turn is influenced by the absorption coefficient and the density. For a cellular plastic material having a density on the order of 10 kg m⁻³, the difference between a 25 and 100 mm thick specimen ranges from 12–15%. This reduces to less than 4% for a density of 48 kg m⁻³. References 23–27 discuss the issue of thickness in more detail.

Mean Temperature. Thermal performance is highly dependent on mean temperature. Figure 4 illustrates the general effect for various typical cellular plastic materials. The inflection in the curve for polyisocyanurate is due to the change of phase of the particular blowing agent from liquid to gas. The position of this inflection depends on the blowing agent used.

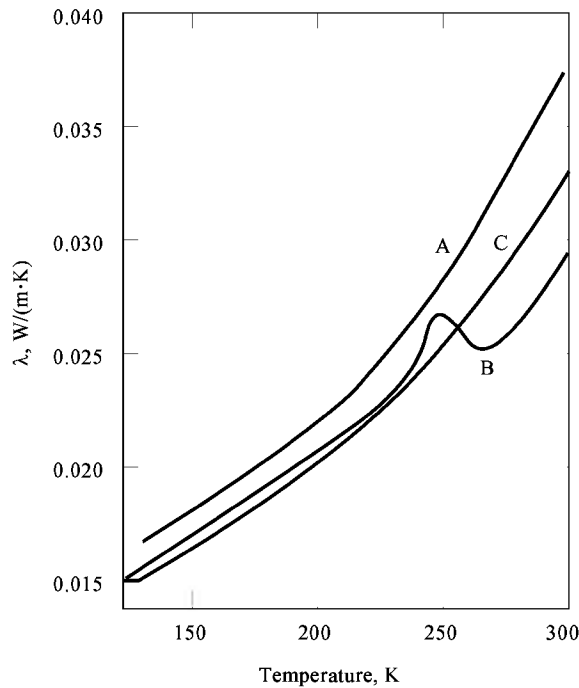


Fig. 4. Thermal conductivity of cellular plastics: A, extruded polystyrene, 32 kg m⁻³; B, polyurethane, 32 kg m⁻³; and C, PVC foam, 32 kg m⁻³.

The thermal conductivities of the most common insulation materials used in construction are shown in Table 2. Values at different mean temperature are necessary for accurate design purposes at representative temperatures encountered during winter or summer. For example, under winter conditions with an outside temperature of -20 to -10°C, the mean temperature is 0–5°C. For summer, mean temperatures in excess of 40°C can be experienced.

Table 2. Thermal Conductivities, W/(m·K)^a

Condition	Mean temperature, °C	XEPS	EPS	Polyisocyanurate (PI)		Phenolic, faced
				Aged	Impermeable facing	
winter	4	0.027	0.032	0.023	0.016	0.016
ambient temperature	24	0.029	0.036	0.025	0.018	0.017
summer	44	0.031	0.040	0.027	0.020	0.018

^a Bulk densities of basic foam, kg m⁻³: EPS, 16; XEPS, 30; aged PI, 32; PI with impermeable facing, 32; and faced phenolic, 45.

Moisture. Absorbed and retained moisture, especially as ice, has a significant effect on the structural and thermal properties of insulation materials. Most closed-cell plastic foams have low permeance properties most notably where natural or bonded low permeance surface skins exist (29,30). Design, building, and construction practices require adequate vapor retarders, skins, coatings, sealants, etc, in order to prevent the presence of moisture. However, moisture vapor cannot be completely excluded, thus the possibility of moisture absorption and retention is always present. The freezing of moisture and rupturing of cells result in permanent reduction of thermal and structural performance.

In standard tests for moisture absorption and water-vapor permeability, material is tested at isothermal temperature conditions. Where such testing has been done, results for extended conditioning periods indicate that it is possible for cellular plastics to absorb and retain large amounts of moisture. Thermal performance can be reduced 20–50% (31–33). However, a similar study (34), simulating roofing environments but under cyclic temperature and humidity conditions resembling those of specific climates, indicates that only small amounts of moisture are retained and thermal performance is reduced by 5% or less. New test methods utilizing specific gradient criteria of temperature and humidity are required to obtain realistic performance characteristics.

Mechanical Properties and Structural Performance. As a result of the manufacturing process, some cellular plastics have an elongated cell shape and thus exhibit anisotropy in mechanical, thermal, and expansion properties (35,36). Efforts are underway to develop manufacturing techniques that reduce such anisotropy and its effects. In general, higher strengths occur for the parallel-to-rise direction than in the perpendicular-to-rise orientation. Properties of these materials show variability due to specimen form and position in the bulk material and to uncertainty in the axes with respect to direction of foam rise. Expanded and molded bead products exhibit little anisotropy.

Strength characteristics are important to consider when selecting materials for particular applications, especially those at low and cryogenic temperatures. Friability is significant in handling and in applications where vibration or movements are involved. Both the mechanical strength and friability depend strongly on density and are also affected by aging and moisture pickup. In general, a mechanical property *MP* is related to density (37):

$$MP = K (\text{density})^\alpha$$

where *K* and *α* are constants depending on the type of foam, orientation, and temperature. Thus for a certain application an optimum density material can be selected with the desired combination of structural and thermal performance. Some typical mechanical properties are given in Table 3.

Table 3. Mechanical Properties of Cellular Plastics, MPa^a

Strength	Polyisocyanurate								Phenolic at 293 K d 35–45 kg m ⁻³
	d 32 kg m ⁻³		d 96 kg m ⁻³		XEPS d 32 kg m ⁻³		EPS at 293 K		
	at 293 K	at 76 K	at 293 K	at 76 K	at 293 K	at 76 K	d = 16 kg m ⁻³	d = 32 kg m ⁻³	
ultimate tensile									
parallel	0.35–0.4	0.4–0.5	1.1	1.6	0.25–0.3	0.20–0.25			0.18–0.25
perpendicular	0.25–0.35	0.3–0.4	1.1	1.7	0.20–0.25	0.15–0.18	0.11–0.14	0.16–0.19	0.13–0.16
tensile modulus									
parallel	10–15	20–30	30	70	25	30			
perpendicular	5–10	10–15	25	60	10	12	1.2–1.5	3.1–3.5	1.28
maximum compressive									

parallel	0.23–0.28	0.31–0.35	0.8	1.4	0.2–0.35	0.3–0.45			0.15–0.20
perpendicular	0.16–0.22	0.17–0.23	0.7	1.3	0.2–0.4	0.2–0.4	0.07–0.1	0.17–0.21	0.10–0.12
compressive modulus									
parallel	7–10	10–13	20	75	12	18			
perpendicular	5–6	5–6	18	60	10	15	1.0–1.4	3.0–3.5	5.7
shear									
parallel	0.16–0.21	0.16–0.24	0.8	1.4	0.2	0.18			0.10–0.12
perpendicular	0.15–0.2	0.12–0.22	0.75	1.3	0.2	0.18	0.12–0.15	0.23–0.25	0.09–0.10

^a To convert MPa to psi, multiply by 145.

Polyurethane, PVC, and extruded polystyrene provide the bulk of the cellular plastics used for low and cryogenic temperature applications. In some cases, eg, the insulation of liquid hydrogen tanks on space systems, foams have been reinforced with continuous glass fibers throughout the matrix. This improves strength without affecting thermal performance significantly.

Flame Resistance. Traditionally, small-scale laboratory flammability tests have been used to initially characterize foams (38). However, these do not reflect the performance of such materials in bulk form. Fire characteristics of thermal insulations for building applications are generally reported in the form of qualitative or semiquantitative results from ASTM E84 or similar tunnel tests (39). Similar larger scale tests are used for aircraft and marine applications.

Although the tunnel test is widely accepted, conditions and orientations involved are not those normally found in installed insulations. New large-scale tests have been developed; the results can be taken to represent actual performance more closely. Such tests include the International Conference of Building Officials (ICBO) and ASTM E603 full-scale room tests, ASTM E108 roofing test, the UL roof deck construction test, the Factory Mutual Calorimeter Test, and both a large- and small-scale corner test.

Flame-spread and smoke-density values, and the less often reported fuel-contributed semiquantitive results of the ASTM E84 test and the limited oxygen index (LOI) laboratory test, are more often used to compare fire performance of cellular plastics. All building codes require that cellular plastics be protected by inner or outer sheathings or be housed in systems all with a specified minimum total fire resistance. Absolute incombustibility cannot be attained in practice and often is not required. The system approach to protecting the more combustible materials affords adequate safety in the buildings by allowing the occupant sufficient time to evacuate before combustion of the protected cellular plastic.

Health and Safety Factors

The long-term effects of CFCs and HCFCs leaking into the environment have been discussed. Combustion where all cellular plastics can evolve smoke containing carbon monoxide and in certain cases cyanide and other toxic gases from various constituents involved in their manufacture is also a consideration.

Urea–formaldehyde use has been greatly restricted because of free formaldehyde (qv) emissions which can cause eye irritation and in some cases serious illness. Some attempts at developing formaldehyde-free urea-based materials are ongoing.

Economic Aspects

In the mid-to-late 1980s, growth estimates of the use of polystyrene and polyurethane cellular plastic insulation materials and products were a healthy 10% per year and greater for phenolic (40,41). The principal application where strongest growth was forecast for these types was for roofing, especially single-membrane systems (42).

From 1991 to 1992, consumption in North America of these forms of cellular plastic (11) rose 5–6% overall from approximately 600 × 10⁶ to 631 × 10⁶ kg. Of this EPS and XEPS each rose only by some 2.5%, and polyurethane by 4%, due solely to its use for appliances. Phenolic and other foams have also contributed but with small growth.

In Europe total consumption of plastic foam insulation for 1992 was 29 × 10⁶ m³ (>12 billion board ft) with little or no growth seen from 1991. All products were expected to grow by an average of 1% due primarily to the continued effects of the economic recession.

Costs of cellular plastic insulations are still higher than those of fibrous and other mass insulation types, but these can often be justified based on overall advantages of combined structural, thermal, and permeance properties. It is difficult to provide a single cost for each material type since there are many different forms of a material-based product available and differing forms of manufacture and application, often in combination with other materials. In the United States, EPS board costs on the order of \$0.12 to \$0.18; XEPS, \$0.25 to \$0.30; and PU, \$0.30 to \$0.35, per board foot (\$0.30 board ft ≈ \$127 m³).

Uses

In addition to building applications, cellular plastics are used in low or cryogenic temperature applications, such as refrigerators and freezers; bulk transportation and storage of foods in refrigerated containers; process industries involving chilled and refrigerant fluids; liquefaction and storage of gases at cryogenic temperatures; pipelines for oil and gas, particularly in Arctic regions; under roads and airport runways in cold regions to resist frost heave; generation and transmission of electricity at low temperatures; marine applications; structures of aircraft and space vehicles and systems; structures and components of missiles and re-entry vehicles; medical and biological sciences, including preservation of blood, organs, and tissues; and electronic equipment where operation at a constant low temperature is essential to diminish electrical noise.

BIBLIOGRAPHY

"Insulation, Thermal" in *ECT* 1st ed., Vol. 7, pp. 927–935, by C. B. Bradley, Johns-Manville Research Center; in *ECT* 2nd ed., Vol. 11, pp. 823–838, by R. H. Neisel and H. F. Remde, Johns-Manville Research & Engineering Center; in *ECT* 3rd ed., Vol. 13, pp. 591–605, by R. H. Neisel and J. D. Verschoor, Johns-Manville Sales Corp.

1. C. J. Benning, *Plastic Foams*, Vol. 1, *The Physics and Chemistry of Product Performance and Process Technology*, John Wiley & Sons, Inc., New York, 1969.
2. A. H. Landrock, *Polyurethane Foams: Technology Properties and Applications*, Report 37, Plastic Technical Evaluation Center, Picatinny Arsenal, Dover, N.J., 1969.
3. *Thermal Insulation Systems—A Survey*, NASA Report SP-5027, NASA, Washington, D.C., 1967.
4. R. N. Miller, C. D. Bailey, R. T. Beal, and J. M. Freeman, *Advances in Cryogenic Engineering*, Vol. 8, Plenum Press, New York, 1963, pp. 417–424; *Ind. Eng. Chem.* **1**(4), 257 (Dec. 1962).
5. F. C. Wilson, *Refrig. Eng.* **65**(4), 57 (1957).
6. *An Assessment of Thermal Insulation Materials and Systems for Building Applications*, DOE Report, BNL-50862 UC-95d, U.S. Dept. of Energy, Washington, D.C., 1978; R. P. Tye and D. L. McElroy, eds., *ASTM STP 718, Thermal Insulation Performance*, American Society for Testing and Materials, Philadelphia, Pa., 1980, pp. 9–26.
7. D. L. Johnston, *Roof Des.* **1**(1), 26 (June 1983).
8. D. L. Johnston, *Roofing Spec.* **11**, 26 (Dec. 1983).

9. C. A. Schutz, *J. Cell. Plast.* **4**(1), 37 (1968).

10. *Roofing/Siding/Insul.* **59**, 79 (Oct. 1982).

11. *Mod. Plast.* 83, 93 (Jan. 1993).

12. C. M. Pelanne, *Therm. Insul.* **1**, 48 (1977).

13. W. C. Turner and J. F. Malloy, *Thermal Insulation Handbook*, R. E. Krieger Publishing Co., Inc., Melbourne, Fla., 1981, pp. 191–275.

14. R. M. E. Diamant, *Steam Heat. Eng.* **33**, 6 (1964).

15. W. R. Strzepek, in E. C. Guyer and D. L. Brownell, eds., *Handbook of Applied Thermal Design*, McGraw-Hill Book Co., New York, 1989, pp. 3–30 to 3–41.

16. *Mod. Plast.*, 1, 8 (Mar. 1993).

17. R. P. Tye, in D. L. McElroy and J. F. Kimpflen, eds., *Insulation Materials Testing and Applications, ASTM STP 1030*, ASTM, Philadelphia, Pa., 1990, pp. 141–155.

18. G. W. Ball, W. G. Healey, and T. B. Partington, *Eur. J. Cell. Plast.* **1**(1), 50 (Jan. 1978).

19. F. J. Norton, *J. Cell. Plast.* **18**, 300 (Sept. Oct. 1982).

20. D. W. Reitz, M. A. Schuetz, and L. R. Glicksman, *J. Cell. Plast.* **20**(2), 104 (1984).

21. F. Sherwood-Rowland, *Chlorofluorocarbons and Depletion of Stratospheric Ozone, Improved Thermal Insulation—Problems and Perspectives*, D. A. Brandreth, ed., Technomic Publishing Co., Inc., Lancaster, Pa., 1991, pp. 5–25.

22. *Protocol on Substances that Deplete the Ozone Layer*, United Nation Environment Programme, Final Act, Montreal, Canada, Sept. 1987.

23. J. Isberg, *The Thermal Conductivity of Polyurethane Foams*, Chalmers University of Technology, Gothenburg, Sweden, 1988.

24. M. T. Bomberg, *J. Thermal Insulation* **13**, 149 (1990).

25. J. R. Booth and J. T. Grimes, *J. Thermal Insulation Build. Envelopes* **15**, 256 (Apr. 1993).

26. B. K. Larkin and S. W. Churchill, *J. AIChE* **5**(4), 467 (1959).

27. T. T. Jones, *Proceedings of the VIth Thermal Conductivity Conference*, NBS Special Publication 302, National Bureau of Standards, U.S. Dept. of Commerce, Washington, D.C., 1967, pp. 737–748.

28. B. Y. Lao and R. E. Skochodopole, *Proceedings of the 4th SPI International Cellular Plastics Conference, Montreal, Nov. 1976*, The Society of the Plastics Industry, New York, 1976, pp. 175–182.

29. F. J. Dechow and K. A. Epstein, in R. P. Tye, ed., *ATM STP 660, Thermal Transmission Measurements of Insulation*, American Society for Testing and Materials, Philadelphia, Pa., 1978, pp. 234–260.

30. G. Ovstaas, S. E. Smith, W. Strzepek, and G. Titley, in F. A. Govan, D. M. Greason, and J. D. McCallister, eds., *ASTM STP 789, Thermal Insulation Materials and Systems for Energy Conservation in the '80s*, American Society for Testing and Materials, Philadelphia, Pa., 1983, pp. 435–454.

31. W. Tobiasson and J. Ricard, *Proceedings of the Fifth Conference on Roofing Technology*, sponsored by NBS and NRCA, Apr. 1979, National Roofing Contractors Association, Chicago, Ill., 1979, pp. 4–16.

32. L. I. Knab, D. R. Jenkins, and R. G. Mathey, "The Effect of Moisture on the Thermal Conductance of Roofing Systems," *NBS Building Science Series 123*, National Bureau of Standards, U.S. Dept. of Commerce, Washington, D.C., Apr. 1980.

33. W. Tobiasson, A. Greatorex, and D. VanPelt, "Wetting of Styrene and Urethane Roof Insulations in the Laboratory and on a Protected Roof Membrane," F. J. Powell and S. L. Matthew, eds., in *Thermal Insulation Materials and Systems STP922*, ASTM, Philadelphia, Pa., 1987.

34. R. P. Tye and C. F. Baker, "Development of Experimental Data on Cellular Plastic Insulations Under Simulated Winter Exposure Conditions," in Ref. 33.

35. L. L. Sparks, in R. P. Tye and D. L. McElroy, eds., *ASTM STP 718, Thermal Insulation Performance*, American Society for Testing and Materials, Philadelphia, Pa., 1980, pp. 431–452; *J. Therm. Insul.* **8**, 198 (Jan. 1985).

36. J. I. DeGisi and T. E. Neet, *J. Appl. Polym. Sci.* **20**, 2011 (1976).

37. R. K. Traeger, *J. Cell. Plast.* **3**(9), 405 (1967).

38. *ASTM Annual Book of Standards*, Vols. 08.01–08.02, American Society for Testing and Materials, Philadelphia, Pa., 1984, Sect. 8.

39. *ASTM Annual Book of Standards*, Vol. 04.06, American Society for Testing and Materials, Philadelphia, Pa. 1984, Sect. 4.

40. *Chem. Eng. News* **62**, 18 (June 25, 1984).

41. *Mod. Plast.* **60**, 72 (Sept. 1983).

42. *Plastic Foam Materials and Roofing Insulation, 1983–1989*, Peter Sherwood Associates, Inc., White Plains, N.Y., Apr. 1984; *Roofing/Siding/Insul.* **61**, 82 (Oct. 1984).

General References

U.S. Residential Insulation Industry, U.S. Dept. of Commerce, Washington, D.C., Aug. 1977, Survey Report.

U.S. Foamed Plastics Markets and Directory, Technomic Publishing Co., Inc., Lancaster, Pa., 1984.

J. Therm. Insul. (now *J. Therm. Insul. Build. Env.*) Technomic Publishing Co., Inc., Lancaster, Pa., 1980–1993.

Ronald P. Tye
Consultant

INSULIN AND OTHER ANTIDIABETIC AGENTS

Diabetes mellitus is a pathologic condition characterized by chronic hyperglycemia (elevated blood glucose) and additional disturbances of carbohydrate, fat, and protein metabolism. In general terms, diabetes mellitus results from too little insulin production or a failure of insulin to function. Normally, blood glucose levels are tightly maintained in the fasting state between 3.6 and 5.3 mmol L (65–95 mg dL) and up to 7.3 mmol L (130 mg dL) following a meal. Following guidelines from the World Health Organization and the American Diabetes Association, the diagnosis of diabetes mellitus should be considered when blood glucose levels are >6.7 mmol L (120 mg dL) following an overnight fast and >10 mmol L (180 mg dL) following meals or an orally ingested glucose [50-99-7], C H₁₂O , challenge (1,2).

There are multiple causes of diabetes. Whereas the molecular bases of some forms of diabetes are well understood, in many cases etiologies are unknown. It is customary to divide diabetes into two main forms: insulin-dependent diabetes mellitus (IDDM), also referred to as Type I or juvenile-onset diabetes, and noninsulin-dependent diabetes mellitus (NIDDM), also called Type II or maturity-onset diabetes (3).

IDDM is the more common form of diabetes in children and young adults. This form of the disease is defined by the presence of such classical diabetes symptoms as extreme thirst, excessive urination, and weight loss in the presence of hyperglycemia. Patients with IDDM are also prone to developing an acute condition called diabetic ketoacidosis; in many cases the diagnosis of IDDM is first made when a patient arrives at the emergency room in this condition. Hyperglycemia and ketoacidosis result directly from the body's inability to utilize circulating glucose and free fatty acids in the absence of insulin. Therefore, dehydration related to hyperglycemia, and additional metabolic abnormalities related to poorly regulated lipolysis (elevated levels of ketones and ketoacids in the blood), lead to this life threatening condition. Precipitating causes of diabetic ketoacidosis include intercurrent infections and other illnesses, serious accidents, and the failure to take adequate doses of insulin. IDDM is caused by autoimmune destruction of the insulin-producing β-cells in the pancreas. In early phases of the disease lymphocytes invade the islets of Langerhans in the pancreas and a variety of autoantibodies can be found in the circulation. The result of this self-directed attack by the immune system is destruction of the pancreatic β-cells and a resultant inability to produce insulin. Additional less common causes of insulin-requiring diabetes which do not result from autoimmune destruction of the islets include pancreatectomy, pancreatic cancers, and pancreatitis.

NIDDM is a much more common disease than IDDM, accounting for about 85–90° of all cases of diabetes mellitus. Whereas NIDDM may be present at any age, the incidence increases dramatically with advanced age; over 10° of the population reaching 70 years of age has NIDDM. Patients with NIDDM do not require insulin treatment to maintain life or prevent the spontaneous occurrence of diabetic ketoacidosis. Therefore, NIDDM is frequently asymptomatic and unrecognized, and diagnosis requires screening for elevations in blood or urinary sugar. Most forms of NIDDM are associated with a family history of the disease, and NIDDM is commonly associated with and exacerbated by obesity. The causes of NIDDM are not well understood and there may be many molecular defects which lead to NIDDM.

Therapy of Diabetes

The goals of diabetes therapy include elimination of the clinical symptoms of hyperglycemia, prevention of diabetic ketoacidosis, normalization of blood sugar values, prevention of long-term sequelae, and restoration of a sense of well-being. Therapeutic regimens are generally tailored to the individual. Patients having IDDM require treatment with insulin. For some patients having NIDDM careful attention to diet and exercise alone may have a profound impact on the disease. Some patients having NIDDM are best managed with insulin as well, whereas others are best treated using oral blood glucose lowering agents. In virtually all cases the requirement for insulin or an oral agent is reduced by proper attention to exercise and diet.

The efficacy of treatment regimens is gauged by the lack of symptoms and monitoring of blood chemistries, including blood glucose and glycosylated hemoglobin levels. The latter tests have revolutionized the approach to diabetic therapy. Blood glucose values can be monitored multiple times each day by the patient, with the goal of maintaining values as close to normal as possible. Various methods available for home blood glucose monitoring, including colorimetric dipsticks and electronic meters, require a small drop of blood; newer, less invasive methods for blood glucose monitoring are under development. Hemoglobin in circulating red blood cells undergoes a nonenzymatic glycosylation to produce what is referred to as glycosylated or glycated hemoglobin, also known as hemoglobin A_{1c}. This reaction takes place continuously over the life of a red blood cell, and its rate is governed by blood glucose concentrations. Therefore, the amount of glycosylated hemoglobin present at a given time reflects an averaged level of blood glucose control over the preceding 6–8 weeks. Taken together, home blood glucose monitoring and determination of glycosylated hemoglobin levels provide extremely useful parameters for assessing the efficacy of treatment.

Prior to the initiation of insulin therapy for IDDM in the 1920s, life expectancy was short owing to severe metabolic derangements and inanition (3). Subsequent to the common use of insulin therapy, life expectancies improved dramatically, but previously unrecognized long-term consequences of diabetes and or its treatment became apparent with the increased longevity. Both IDDM and NIDDM lead to tissue damaging complications, which may be divided into microvascular and macrovascular. Microvascular refers to small blood vessels and the resulting complications include diseases of the eyes (retinopathy), kidneys (nephropathy), and nerves (neuropathy). Diabetic retinopathy is now the greatest cause of blindness in the United States among persons over 21 years of age. Proliferative retinopathy in particular leads to blindness, but if treated early by laser photocoagulation therapy, visual loss can be significantly reduced (see LASERS). Diabetic nephropathy may ultimately lead to renal failure requiring dialysis or kidney transplant. Neuropathies associated with diabetes can cause pain, burning or loss of sensation, loss of function (eg, impotence), diarrhea, and postural hypotension. Macrovascular complications arise from atherosclerosis, which results in reduced blood flow. Complications include angina and myocardial infarctions, stroke, and vascular insufficiency of the lower extremities leading to amputations.

A principal question in the therapy of diabetes has been whether normalization of blood glucose levels would reduce the incidence of these serious long-term complications of diabetes. Although tight glycemic control is desirable, it is often difficult to achieve and can be accompanied by the potentially life-threatening side effect of tight control, hypoglycemia. After years of debate a landmark study, the Diabetes Control and Complications Trial (DCCT), was designed. A total of 1441 patients with IDDM were recruited to the study from 29 centers between 1983–1989. The trial was terminated in 1993 when it became clear that tight glycemic control significantly reduced the risk of developing retinopathy, nephropathy, and neuropathy (4).

Insulin

Insulin [9004-10-8] is a peptide hormone produced in the islets of Langerhans within the pancreas, which acts as the principal regulator of glucose homeostasis. Under normal physiological conditions the β-cells within pancreatic islets secrete insulin into the bloodstream following nutrient ingestion. The insulin is then carried in the blood to targeted tissues, all of which have insulin receptors on their cellular surfaces. Tissues of insulin action include the liver, muscle, and fat. Insulin binding to the insulin receptors on these cells induces a series of intracellular events which culminate in the increased cellular uptake of circulating glucose and other nutrients as well as their storage as glycogen and fat, and increases in gene expression and protein synthesis.

The classic experiments of Von Mering and Minkowski in 1889 first implicated the pancreas in regulating blood glucose levels: removal of a dog's pancreas led directly to the development of hyperglycemia. Then in the early 1920s it was shown that an internal secretion of the pancreas could be isolated

and used to lower the elevated blood glucose levels of pancreatectomized dogs. Subsequently pancreatic extracts were used to treat patients with diabetes as well (5). Frederick Banting and John Macleod were awarded the 1923 Nobel Prize in Medicine for these discoveries.

Chemistry. Although insulin was recognized to be a protein shortly after its discovery, its primary structure was not elucidated until the 1950s (6). Insulin was the first protein to have its entire primary sequence determined, and for this achievement Frederic Sanger received the 1959 Nobel Prize in Chemistry. All known insulins are composed of two polypeptide chains linked to one another by disulfide bonds; the structures of human, pig, and beef insulins are compared in Figure 1. The A-chain contains 21 amino acids (qv); the B-chain has 30 amino acids. These two peptide chains are covalently linked to one another by two cystine disulfides, one between CysA7 and CysB7, and the other between CysA20 and CysB19. An additional intrachain disulfide connects cysteines A6 and A11.

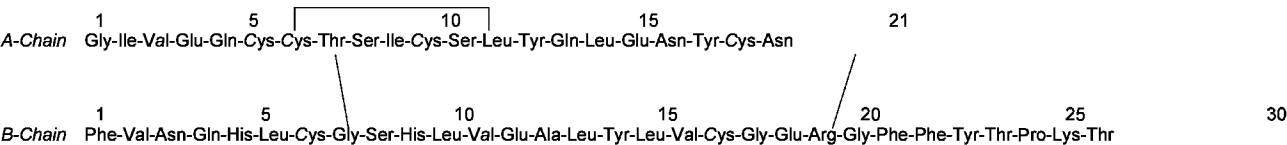


Fig. 1. Amino acid sequence for the A- and B-chains of human insulin [11061-68-0] where solid lines denote disulfide bonds. Porcine insulin [12584-58-6] differs by one amino acid in the B-chain where alanine replaces threonine at position 30. Bovine insulin [11070-73-8] differs by three amino acids. In the A-chain alanine replaces the threonine at position 8 and valine replaces the isoleucine at position 10. In the B-chain there is an alanine at position 30.

For many years patients with diabetes were treated with insulins that had been isolated from the pancreases of pigs and cows. The primary sequences of these insulins are closely related to the sequence of human insulin. There is only a single difference between the sequences of human and porcine insulins: human insulin has a threonine at position B30, and porcine insulin has an alanine. Bovine insulin differs from human insulin at three positions. There is an alanine at the B30 position, an alanine at A8, and a valine at A10. These conservative changes in primary sequence have no apparent effect on biologic activity, although there are slight differences in solubility (7).

In crystal structures the two chains of insulin form highly ordered globular structures (8). Two main structural types form depending on crystallization conditions. In both structures the A-chains form two α -helical segments, from residues A1–A8 and A13–A19, which are connected by a turn. In the structure referred to as the T-state, the B-chain contains two regions of extended chain, B1–B8 and B21–B30, connected by an α -helix from B9–B19. In the R-state structure, the B-chain helix extends from B1–B19. The crystallographic T-state structure best matches the solution structure of insulin determined by nmr (9), although the R-state can be induced in solution under the appropriate conditions. The surface of insulin which interacts with the insulin receptor includes the N- and C-termini of the A-chain and the C-terminus of the B-chain.

Preparation of Insulins. Until the early 1980s insulin for therapeutic purposes was produced almost exclusively by extraction from beef and pork pancreases. Between 100 and 400 mg of insulin can be obtained from each kg of pancreatic tissue, and it has been estimated that there would be sufficient supplies of animal insulin to meet the requirements of diabetic patients into the twenty-first century (2). Through modern purification procedures animal insulins can be prepared in essentially pure form, which eliminates the possibility of developing antibodies against impurities in the insulin preparations. However, patients treated with purified insulins still develop antibodies to insulin, suggesting that differences in the primary structures of these insulins might stimulate antibody production. Therefore, enzymatic and biosynthetic methods have been developed for the preparation of therapeutic insulin identical to human insulin.

Isolation of Animal Insulins. The underlying procedure for isolating insulin from pancreatic tissue has remained nearly unchanged since the 1930s (7). Frozen pancreases are extracted in acidified aqueous ethanol to solubilize the insulin and inactivate exocrine proteases. Following a neutralization step, the pH is readjusted to between 3 and 4 (near the isoelectric point of insulin), 2–3 M sodium chloride is added, and insulin precipitates. Salt cake insulin is then redissolved in acid and crystallized in the presence of zinc. After two crystallization steps insulin of 80–90% purity is obtained. This product contains substantial amounts of proinsulin and related conversion intermediates, arginyl- and ethyl-insulins, covalent insulin dimers, and monodesamido insulin. Because these impurities are potentially deleterious, most insulin preparations used to treat diabetic patients are further purified by gel-filtration, ion-exchange, and or reversed-phase liquid chromatography. Highly purified insulin preparations for treating diabetic patients including single-component (SC) insulin (Eli Lilly and Co.) and monocomponent (MC) insulin (Novo Nordisk) contain <1 ppm of impurities.

Preparation of Human Insulin. Porcine insulin can be converted to the human insulin sequence by an enzyme-catalyzed transeptidation reaction (10,11). Under appropriate conditions trypsin acts preferentially at LysB29 rather than ArgB22 to yield a covalent des[B30]insulin trypsin complex (acyl-enzyme intermediate). In the presence of high concentrations of organic co-solvents and the *t*-butyl ester of threonine, transeptidation predominates over hydrolysis to yield the *t*-butyl ester of human insulin. Following appropriate purification steps and acidolytic removal of the ester, human insulin suitable for treating patients is obtained.

Advances in recombinant deoxyribonucleic acid (DNA) technology make possible the commercial production of human insulin for therapeutic purposes from protein products produced either in bacteria or yeast (see GENETIC ENGINEERING). In fact, the biosynthetic production of human insulin can be viewed as the prototype for the biotechnology industry. Over 1000 kg of purified human insulin is made and consumed each year in the United States alone. Three basic strategies have been used (12–14). In the first, A- and B-chains were synthesized independently, the cysteine residues were converted to the respective S-sulfonates to ease handling and purification, and following mild reduction and air oxidation, intact human insulin was obtained. In the second approach, full-length proinsulin [9035-68-1] was produced in *E. coli* following a similar strategy, but the post-fermentation chemistry was simplified because proper chain combination is directed by the C-peptide. The proinsulin product was cleaved enzymatically to yield intact human insulin. In the third approach, a shortened miniproinsulin having a three residue C-peptide was produced in yeast (*S. cerevisiae*). The miniproinsulin was secreted by the yeast into the media as a properly processed and folded product. After purification steps and trypsin-catalyzed transeptidation and acidolysis, intact human insulin was obtained. Human insulin produced in *E. coli* is available from Eli Lilly and Co., whereas human insulin produced in yeast is marketed by the Novo Nordisk Co. Nearly all patients initiating insulin therapy receive human insulin and many of those that have previously received bovine and or porcine insulin have been switched to recombinant human insulin.

Therapeutic Insulin Preparations. Insulin preparations for therapeutic use differ in time of onset, duration of action, purity, and species of origin (15). The concentration of all nonprescription insulins available in the United States is U100 (100 units mL). A unit is the amount of insulin required to reduce the blood glucose of a fasting rabbit to 2.5 mmol L (45 mg dL). There are generally 24–30 units mg of purified insulin. A U500 insulin preparation, Regular (Concentrated) Iletin II, is available by prescription for the treatment of severe insulin resistance. Insulin must be given by hypodermic injection or infusion pump because the hormone is destroyed in the gastrointestinal tract. Individuals having diabetes are trained to inject themselves. For this purpose a special syringe measuring the dosage of insulin directly in units is employed. Insulins may be divided into rapid-, intermediate-, and long-acting preparations depending on the rapidity of onset and duration of action (Table 1).

Table 1. Characteristics of Insulin Preparations

Composition	Preparation ^a	Action profile, h ^b			Insulin species ^c
		Onset	Peak	Duration	
		<i>Short-acting</i>			
insulin solution unbuffered, regular		0.5	2–5	6–8	H,P,B P
phosphate buffer, buffered regular		0.5	2–5	6–8	H
		<i>Intermediate-acting</i>			

protamine zinc suspension, phosphate buffer, NPH	1–2	4–12	18–26	H,P,B,B P
amorphous and crystalline suspension, acetate buffer, lente	1–3	6–15	18–26	H,P,B,B P
NPH 70° o, regular 30° o, isophane regular	0.5	2–12	24	H,P
NPH 50° o, regular 50° o	0.5	2–12	24	H
<i>Long-acting</i>				
crystalline suspension, acetate buffer, ultralente	4–6	8–30	24–36	H,B

^a See text.
^b Times are averages and can vary markedly between patients, insulin species, injection site, etc.
^c H human insulin; P porcine insulin; B bovine insulin; B P bovine porcine mixture.

Rapid-Acting Insulin Preparations.

Insulin Injection. Regular insulin, crystalline zinc insulin, has both a relatively rapid onset and a short duration of action. It may be given intravenously and intramuscularly, as well as subcutaneously. It is substantially free from turbidity and insoluble matter, and contains 0.1–0.25% wt vol of either phenol or cresol and 1.4–1.8% wt vol of glycerol. Its pH is 2.5–3.5 for acidified injection and 7.0–7.8 for neutral injection. The unpurified is known as Regular Insulin or Regular Iletin I; the purified as Purified Pork Insulin, Regular Iletin II, or Velosulin. Regular insulin [9004-10-8] is widely used to supplement intermediate- and long-action preparations, and, when buffered, it is the insulin used with infusion pumps. Insulin mixtures provide more flexibility in delivering appropriate amounts of insulin at the time of food intake. Regular insulin is the preparation of choice in unstable diabetes when complications such as infection, shock, or surgical trauma occur. It may be administered intravenously to treat ketoacidosis or during surgery.

Regular Human Insulin Injection. Known as Humulin R, Novolin R, or Velosulin Human, this rapid-acting form of human insulin [11061-68-0] is produced by recombinant DNA techniques (biosynthetic) or enzymatic conversion (semisynthetic) of porcine insulin. It may be administered subcutaneously, intravenously, intramuscularly, or through an infusion pump. Therapeutically, this preparation is probably equivalent to purified porcine insulin injection. Human regular insulins are absorbed faster than the corresponding purified porcine product in some patients. Although the peak serum concentration of human insulin injection after subcutaneous administration is slightly higher than that of purified porcine insulin injection, the times to peak concentration and overall bioavailability are similar and control of blood glucose appears to be equivalent. No differences are apparent in binding, actions, metabolism, or potency.

Intermediate-Acting Preparations.

Isophane Insulin Suspension. Isophane insulin [8052-74-2] is an intermediate-acting preparation. The unpurified is known as NPH Iletin I or NPH Insulin; the purified as Insulatard NPH, NPH Iletin II, or NPH Purified Pork Insulin. Absorption is delayed because the insulin is conjugated with protamine in a complex of reduced isoelectric solubility, such that the solid phase of the suspension consists of crystals composed of insulin, protamine, and zinc. The protamine sulfate [9009-65-8] is prepared from the sperm or from the mature testes of fish. This preparation is useful in all forms of diabetes except the initial treatment of diabetic ketoacidosis or in emergencies. Isophane insulin is never given intravenously. Hypoglycemic reactions in mid-to-late afternoon may be less obvious in onset, more prolonged, and more frequent than for rapid-acting preparations because of the prolonged effect of the dose.

Isophane insulin is a white suspension of rod-shaped crystals approximately 30-µm long, and free from large aggregates of crystals after being subjected to moderate agitation. It contains either 1.4–1.8° o glycerol, 0.15–0.17° o *meta*-cresol, and 0.06–0.07° o phenol on a wt vol basis, or 1.4–1.8° o glycerol and 0.20–0.25° o phenol (wt vol), at a pH of 7.1–7.4. It also contains 0.15–0.25° o (wt vol) of sodium phosphate, 0.01–0.04 mg of zinc, and 0.3–0.6 mg of protamine for each USP insulin unit. The insoluble matter in the suspension is crystalline and contains not more than traces of amorphous material.

NPH Isophane Human Insulin Suspension. NPH isophane insulin, also called Humulin N, Insulatard NPH Human, or Novolin N is an intermediate-acting form of human insulin produced by recombinant DNA techniques. Mixtures Humulin 70 30 and Novolin 70 30 contain 70° o NPH isophane and 30° o regular, whereas Humulin 50 50 contains 50° o NPH isophane and 50° o regular. It is administered subcutaneously and should not be given intravenously. Absorption is delayed because the insulin is conjugated with protamine in a complex of reduced isoelectric solubility. Therapeutically, this preparation is probably comparable to purified porcine NPH insulin. However, human NPH insulin may have a slightly shorter duration of action than comparable purified porcine products.

Insulin Zinc Suspension. Lente insulin [8049-62-5], where the unpurified is called Lente Iletin I or Lente Insulin, and the purified, Lente Iletin II or Lente Purified Pork Insulin, is an intermediate-acting preparation composed of 30° o prompt insulin zinc suspension (semilente insulin) and 70° o extended insulin zinc suspension (ultralente insulin) (see Table 1). Insulin zinc suspension or isophane insulin (usually in combination with regular insulin) is often used for previously untreated diabetic patients who require insulin. Insulin zinc suspension is not a suitable substitute for regular insulin in emergencies because of its delayed onset of action. Insulins of the lente series can be mixed in any proportion to obtain the desired dose and modified activity. The advantage of zinc insulin suspension is its freedom from foreign proteins, eg, globin or protamine, to which some patients are sensitive.

Lente insulin is an almost colorless suspension of a mixture of characteristic crystals predominantly 10–40 µm in maximum dimension, and many particles that have no uniform shape and do not exceed 2 µm in maximum dimension. On a wt vol basis, it contains 0.15–0.17° o sodium acetate, 0.65–0.75° o sodium chloride, 0.09–0.11° o methylparaben, and 0.20–0.25 mg of zinc of which 40–65° o is in the supernatant liquid. Its pH is 7.1–7.5.

Human Insulin Zinc Suspension. This insulin, Humulin L or Novolin L, is an intermediate-acting form of human insulin produced by recombinant DNA techniques. It is administered subcutaneously and should not be given intravenously. Therapeutically, this preparation is probably comparable to purified porcine insulin zinc suspension. However, human insulin may have a slightly shorter duration of action than comparable purified pork products.

Long-Acting Insulin Preparations.

Extended Insulin Zinc Suspension. Ultralente insulin is an unpurified, sterile suspension of insulin in buffered water for injection, modified by the addition of zinc chloride in a manner such that the solid phase of the suspension is crystalline. The actions, indications, and potential for hypoglycemic reactions of this long-acting preparation resemble those of protamine zinc insulin. Like prompt insulin zinc suspension (semilente insulin), this form contains no modifying protein to which patients may be sensitive. Because of its long duration of action, this insulin preparation has limited usefulness when given alone. It is usually administered in combination with a shorter acting form. In slightly reduced doses, it may be combined with insulin zinc suspension (lente insulin) when blood glucose levels are not adequately controlled during the day. Insulins of the lente series can be mixed in any proportion to obtain the desired dose and modified activity. Extended zinc insulin suspension is not suitable for use in emergencies because of its delayed onset of action.

Ultralente Insulin is an almost colorless suspension of a mixture of characteristic crystals, the maximum dimension of which is predominantly 10–40 µm. It contains, for each 100 USP units of insulin, 0.20–0.25 mg of zinc (of which 40–65° o is in the supernatant liquid), and not more than 0.70 mg of nitrogen. It also contains on a wt vol basis 0.15–0.17° o sodium acetate, 0.65–0.75° o sodium chloride, and 0.09–0.11° o methylparaben. The purified form is available as Ultralente Purified Beef Insulin.

Human Extended Insulin Zinc Suspension. Ultralente Humulin U is a long-acting form of human insulin produced by recombinant DNA techniques. It is administered subcutaneously and should not be given intravenously. The time course of this preparation is similar for onset of activity but shorter for maximum activity and duration of action compared with ultralente preparations of animal origin. Insulins of the lente series can be mixed in any proportion to obtain the desired dose and modified activity.

Oral Hypoglycemic Agents

Three classes of oral therapeutic agent are available for treating patients with diabetes mellitus (NIDDM): the arylsulfonylureas (known simply as sulfonylureas), biguanides, and α -glycosidase inhibitors. Since 1977, only the sulfonylureas have been approved for use in the United States, although the other classes are used elsewhere.

Sulfonylureas. The hypoglycemic effect of sulfonylureas was first noted in the early 1940s when several patients died in hypoglycemic coma after testing glyprothiazole, a synthetic sulfonamide used to treat typhoid. Chemical modifications which enhanced activity and lowered toxicity led to the development of the first-generation sulfonylureas. Carbutamide [339-43-5], C₁₁H₁₇N₃O₃S, the first commercial sulfonylurea, came onto the European market in 1955 as a blood-sugar lowering agent, but was later withdrawn owing to toxicity. Tolbutamide [64-77-7], C₁₂H₁₈N₂O₃S, was the first sulfonylurea to be used widely in the treatment of NIDDM.

Structures of sulfonylureas commonly used as oral hypoglycemic agents are shown in Table 2. The central moiety of these compounds confers the hypoglycemic activity. Substituents on the phenyl ring (R) and the urea groups (R') affect differences in potency, duration of action, and toxicity. Additional chemical modifications of substituents lead to the development of the second-generation sulfonylureas. The mechanism of action of all sulfonylureas are similar, although second-generation agents have much higher intrinsic activity.

Table 2. Sulfonylureas Used as Oral Hypoglycemic Agents

Compound	CAS Registry Number	Molecular formula	R R'
<i>First generation</i>			
tolbutamide	[64-77-7]	C ₁₂ H ₁₈ N ₂ O ₃ S	
chlorpropamide	[94-20-2]	C ₁₀ H ₁₃ ClN ₂ O ₃ S	
tolazamide	[1156-19-0]	C ₁₄ H ₂₁ N ₃ O ₃ S	
acetohexamide	[968-81-0]	C ₁₅ H ₂₀ N ₂ O ₄ S	
<i>Second generation</i>			
glyburide (glibenclamide)	[10238-21-8]	C ₂₃ H ₂₈ ClN ₃ O ₅ S	
glipizide	[29094-61-9]	C ₂₁ H ₂₇ N ₅ O ₄ S	
gliclazide	[21187-98-4]	C ₁₅ H ₂₁ N ₃ O ₃ S	

The absorption of sulfonylureas from the upper gastrointestinal tract is fairly rapid and complete. The agents are transported in the blood as protein-bound complexes. As they are released from protein-binding sites, the free (unbound) form becomes available for diffusion into tissues and to sites of action. Specific receptors are present on pancreatic islet β -cell surfaces which bind sulfonylureas with high affinity. Binding of sulfonylureas to these receptors appears to be coupled to an ATP-sensitive K⁺ channel to stimulate insulin secretion. These agents may also potentiate insulin-stimulated glucose transport in adipose tissue and skeletal muscle.

Metabolism of sulfonylureas occurs mainly in the liver, either to inactive (tolbutamide, tolazamide, glipizide, glyburide) or active (acetohexamide, chlorpropamide) compounds that are excreted mainly in the urine. Metabolism of chlorpropamide is incomplete, and about 20% is excreted unchanged; the metabolites of acetohexamide are more potent than the parent compound. Therefore, these drugs can be a problem in patients having impaired renal function. All sulfonylureas are metabolized in the liver and should be avoided in patients having significant hepatic dysfunction. Metabolites of glyburide on the other hand are secreted in the bile, as well as the urine, making glyburide the agent of choice in patients experiencing compromised renal function.

Drug Selection. Studies comparing the effectiveness of sulfonylurea agents have shown that chlorpropamide, tolazamide, glyburide, and glipizide are equally effective; tolbutamide and acetohexamide are less effective. If blood glucose is not controlled in a patient receiving the maximal dose of one of the four most effective sulfonylureas, substituting the maximal dose of another of these agents restores diabetic control in only a small (<10%) percentage of patients.

The second-generation drugs are nearly equivalent therapeutically, but some differences exist. Although both glipizide and glyburide often can be administered once daily, the former has a shorter serum half-life and usually requires twice-daily administration. Its concentration in the serum rises higher and more rapidly and is accompanied by greater insulin secretion initially. Glyburide is more likely to increase sensitivity to insulin. Glipizide may reduce post-prandial blood glucose more effectively, whereas glyburide may be effective in decreasing fasting blood glucose levels and enhancing basal insulin secretion.

Relative potency alone does not determine drug selection because maximal effectiveness is similar for all agents. A single daily dose of any sulfonylurea, except tolbutamide, is sometimes adequate to control blood glucose in NIDDM patients.

Acetohexamide. Acetohexamide or 1-[(*p*-acetylphenyl)sulfonyl]-3-cyclohex-ylurea, mol wt 324.42, is a white, practically odorless crystalline powder that has the trade name Dymelor. It is practically insoluble in water and in ether, soluble in pyridine and in dilute solutions of alkali hydroxides, and slightly soluble in alcohol and in chloroform.

The incidence of untoward effects is low and the reactions are reversible when acetoexamide is discontinued. Relatively severe hypoglycemic reactions have been observed occasionally in patients given large doses for a prolonged period without close observation. It is contraindicated in patients having hyperglycemia and glycosuria owing to primary renal disease, and in those who are hypersensitive to sulfonylurea compounds.

Chlorpropamide. Chlorpropamide (1-[(*p*-chlorophenyl)sulfonyl]-3-propylurea), mol wt 276.75, is a white, crystalline powder, having a slight odor, mp 127–129°C. It is sold as Diabinese, is soluble in water at pH 6 (2.2 mg mL), and practically insoluble at pH 7.3, soluble in alcohol, and sparingly soluble in chloroform, ether, and benzene.

Untoward reactions have been reported more frequently with chlorpropamide than with other sulfonylureas, and the drug should be used with caution. In a few older patients, hypoglycemic reactions have been severe. Water retention with hyponatremia can be life-threatening in patients having a tendency to retain water, eg, those with congestive heart failure or hepatic cirrhosis. Elderly patients and those taking thiazide diuretics may be more likely to develop this complication. This drug should not be used in patients with renal insufficiency, because the duration of action is greatly prolonged. This drug generally is not given to elderly patients.

Facial flushing after ingestion of alcohol occurs in up to one-third of patients taking chlorpropamide. The mechanism, like that of the disulfiram reaction, probably involves inhibition of the oxidation of acetaldehyde, a metabolite of ethanol. The plasma concentration of chlorpropamide may be correlated with chlorpropamide–alcohol flushing.

Glipizide. Glipizide (1-cyclohexyl-3[[*p*-[2-(5-methylpyrazinecarboxamido)ethyl]phenyl]sulfonyl]urea), mol wt 445.55, forms crystals from ethanol, mp 208–209°C. It is known commercially as Glucotrol.

Glipizide is relatively free of serious adverse effects and only approximately 1.5% of patients discontinue this drug because of adverse reactions. Gastrointestinal disturbances are most common (incidence 1.7–3.7%); skin rashes occur in up to 1.4% of patients.

Glyburide. Glyburide or 1-[[*p*-[2-(5-chloro-*o*-anisamido)ethyl]phenyl]sulfonyl]-3-cyclohexylurea, mol wt 494.00, forms crystals from methanol, mp 169–170°C. Its pK_a is 5.3 and it is sparingly soluble in water and soluble in the usual organic solvents.

The incidence of serious side effects with glyburide, sold as DiaBeta and Micronase, is low. Gastrointestinal disturbances develop in 1.8% of patients. Skin rashes occur in 1.5% of patients and may disappear with continued use.

Tolazamide. Tolazimide (1-(hexahydro-1*H*-azepin-1-yl)-3-(*p*-tolysulfonyl)urea), mol wt 311.40, is a white to off-white crystalline powder, odorless or having a slight odor, mp 170–173°C, with a pK_a of 3.6 at 25°C and 5.68 at 37.5°C. It is very slightly soluble in water, freely soluble in chloroform, soluble in acetone, and slightly soluble in alcohol. The trade name is Tolinase.

Generally, the untoward effects associated with tolazamide are the same as those noted with the other sulfonylureas; the incidence is low and reactions are reversible when tolazamide is discontinued. Hypoglycemia has been reported occasionally.

Tolbutamide. Tolbutamide (1-butyl-3-(*p*-tolylsulfonyl)urea), with mol wt 270.35, is known as Orinase. It is a white to off-white practically odorless crystalline powder having a slightly bitter taste, mp 126–132°C. It is practically insoluble in water, and soluble in alcohol and chloroform. The toxicity of tolbutamide appears to be low, and reactions are similar to those observed with other sulfonylureas.

Tolbutamide Sodium USP. Orinase Diagnostic [473-41-6] (*N*-[(butylamine)carbonyl]-4-methylbenzenesulfonamide, monosodium salt), mol wt 292.33, is a white to off-white practically odorless crystalline powder having a slightly bitter taste. It is freely soluble in water, soluble in alcohol and chloroform, and very slightly soluble in ether and can be prepared by dissolving tolbutamide in aqueous NaOH.

Because of its water solubility, tolbutamide sodium may be given intravenously (1 g of tolbutamide equivalent over a period of 2–3 min). By this route, its rapid onset of action lends itself to the diagnosis of diabetes mellitus in persons in whom the usual studies are equivocal. The normal person responds with a more rapid and intense drop in blood glucose than does the diabetic, especially during the first hour after injection. Persons having pancreatic insulinoma respond with a prolonged hypoglycemia, so that the drug may also be used diagnostically when that condition is suspected. Tolbutamide sodium is an irritant, and thrombophlebitis and thrombosis of the vein occur in ca 1–2% recipients.

Biguanides. Biguanides, which are guanidine derivatives, were introduced into clinical use for the treatment of hyperglycemia in patients with Type II diabetes mellitus in the 1950s. Three biguanides were available initially: phenformin, metformin, and butformin. Phenformin, an investigational drug widely prescribed in the United States, and butformin have both been banned for clinical use because of a significant incidence of associated lactic acidosis. Metformin, introduced in France in 1959, continues to be used worldwide and is undergoing clinical investigation in the United States. Because of a different mechanism of action, biguanides might be used instead of or in combination with sulfonylureas.

Phenformin. Phenformin hydrochloride [834-28-6] (1-phenethylbiguanide, *N*-(2-phenylethyl)imidodicarbonimidic diamide), is a white to off-white odorless crystalline power having a bitter taste. The melting point is 175–178°C. It is freely soluble in water and alcohol, and practically insoluble in chloroform, ether, and hexane. Its pH in solution is 6.0–7.0.

Metformin. Metformin [657-24-9] (1,1-dimethylbiguanide), mol wt 129.17, forms crystals from propanol, mp 218–220°C, and is soluble in water and 95% ethanol, but practically insoluble in ether and chloroform. Metformin, an investigational drug in the United States, does not increase basal or meal-stimulated insulin secretion. It lowers blood glucose levels in hyperglycemic patients with Type II diabetes but has no effect on blood glucose levels in normal subjects. It does not cause hypoglycemia. Successful metformin therapy usually is associated with no or some weight loss.

Clinical studies have shown that metformin is as effective as chlorpropamide, tolbutamide, or gliclazide in lowering fasting and post-prandial hyperglycemia in patients with newly diagnosed Type II diabetes. However, metformin treatment produces a modest mean weight loss of 1–2 kg as compared with a mean weight gain of 1.5–5.0 kg produced by sulfonylurea therapy. Therapy using metformin results in better glycemic control in about 80% of both obese and nonobese patients having newly diagnosed Type II diabetes. Reported rates of primary failure with metformin are about 12%, and those of secondary failure are about 5%.

Inhibitors of α -Glucosidase. Acarbose [56180-94-0] (*O*-4,6-dideoxy-4-[[[1*S*-(1 α ,5 β ,6 α)]-4,5,6-trihydroxy-3-(hydroxymethyl)-2-cyclohexen-1-yl]amino]- α -D-glucopyranosyl-(1 $\rightarrow$ 4)-*O*- α -D-glucopyranosyl-(1 $\rightarrow$ 4)-D-glucose), with mol wt 645.63, is an investigational drug. It is a pseudotetrasaccharide isolated from strains of *Actinoplanes* and contains an unsaturated cyclitol moiety.

Acarbose is a nonabsorbable α -glucosidase inhibitor which blocks the digestion of starch, sucrose, and maltose. The digestion of complex carbohydrates is delayed and occurs throughout the small intestine rather than in the upper part of the jejunum. Absorption of glucose and other monosaccharides is not affected. Acarbose is administered orally three times a day and chewed with the first mouthful of food.

Miscellaneous Agents

Diazoxide [364-98-7] (3-methyl-7-chloro-1,2,4-benzothiadiazine 1,1-dioxide) is a nondiuretic thiazide used for its hyperglycemic actions when given orally (Proglycem) and for its antihypertensive effects when given intravenously (Hyperstat I.V.) (see CARDIOVASCULAR AGENTS). Diazoxide produces a prompt dose-related increase in blood glucose by directly inhibiting insulin secretion and, possibly, by stimulating epinephrine secretion by the adrenal medulla (see EPINEPHRINE AND NOREPINEPHRINE). Diazoxide is used to counteract hyperinsulinism in conditions such as insulinoma. It is not indicated in the treatment of functional hypoglycemia.

Glucagon [16941-32-5] is a polypeptide produced by the alpha cells of the pancreas. Like insulin, its normal function appears to be to control the homeostasis of glucose, amino acids, and possibly free fatty acids. However, its potent glycogenolytic and gluconeogenic effects are opposite to those of insulin and form the basis for glucagon's clinical usefulness which is the treatment of severe hypoglycemia in patients with diabetes. Glucagon must be administered parenterally. It often is given to a patient by a family member or other caretaker when a severe hypoglycemic episode occurs and the patient is unable to ingest sugar or simple carbohydrates or is unconscious. Glucagon increases the blood glucose concentration by mobilizing hepatic glycogen and thus is effective only when hepatic glycogen is available. Patients having reduced glycogen stores, eg, starvation, adrenal insufficiency, or alcoholic hypoglycemia, cannot respond to glucagon.

Glucagon also has been used to diagnose insulinoma and pheochromocytoma. For the former, the rise in plasma insulin concentration following

intravenous glucagon may be diagnostic.

BIBLIOGRAPHY

"Insulin" in *ECT* 1st ed., Vol. 7, pp. 935–941, by H. F. Jensen, Army Medical Research Laboratory, Fort Knox; *ECT* 2nd ed., Vol. 11, pp. 838–846, by P. Turi, Sandoz, Inc.; "Insulin and Other Antidiabetic Agents" in *ECT* 3rd ed., Vol. 13, pp. 605–620, by G. F. Tutwiler, McNeil Laboratories.

1. National Diabetes Data Group, *Diabetes* **28**, 1039 (1979).
2. WHO Expert Committee on Diabetes mellitus, *WHO Tech Rep Ser.* **646**(2) (1980).
3. C. R. Kahn and G. C. Weir, eds., *Joslin's Diabetes Mellitus*, 13th ed., Lea & Febiger, Philadelphia, Pa., 1994.
4. The Diabetes Control and Complications Trial Research Group, *New Engl. J. Med.* **329**, 977 (1993).
5. M. Bliss, *The Discovery of Insulin*, University of Chicago Press, Chicago, 1982.
6. F. Sanger, *Science* **129**, 1340 (1959).
7. J. Brange, *Galenics of Insulin: The Physico-Chemical and Pharmaceutical Aspects of Insulin and Insulin Preparations*, Springer-Verlag, Berlin, 1987.
8. E. N. Baker and co-workers, *Philos. Trans. Res. Soc. London [Biol]* **319**, 369 (1993).
9. Q. X. Hua, S. E. Shoelson, M. Kochoyan, and M. A. Weiss, *Nature* **354**, 238 (1991).
10. K. Morihara, T. Oka, and H. Tsuzuki, *Nature* **280**, 412 (1979).
11. J. Markussen, *Human Insulin by Tryptic Transpeptidation of Porcine Insulin and Biosynthetic Precursors*, MTP Press, Lancaster, U.K., 1987.
12. R. E. Chance and co-workers, in D. H. Rich and E. Gross, eds., *Peptides: Synthesis–Structure–Function*, Pierce Chemical Co. Press, Rockford, Ill., 1981, p. 721.
13. B. H. Frank, J. M. Pettee, R. E. Zimmerman, and P. J. Burck, in Ref. 12, p. 729.
14. J. Markussen and co-workers, in D. Theodoropoulos, ed., *Peptides 1986: Proceedings of the 19th European Peptide Symposium*, Walter de Gruyter, Berlin, 1986, p. 189.
15. *Annual 1994 AMA Drug Evaluations*, American Medical Association, John Wiley & Sons, Inc., New York, 1994, p. 1019.

Steven E. Shoelson
Joslin Diabetes Center, Harvard Medical School

INTEGRATED CIRCUITS

Since its inception in the 1960s, the electronics industry has driven phenomenological changes in the economies of the world. The fundamental cornerstone of the development of the electronics age has been the concurrent decrease in cost associated with designing and manufacturing an increasingly sophisticated and miniaturized unit of integrated circuitry. The basic building block of an integrated circuit (IC) is a device or component which is a specific combination of conducting, semiconducting, and nonconducting layers that performs a single electrical function on an IC chip. Many devices are connected (or integrated) into a complete integrated circuit, or chip. Thousands of chips are formed simultaneously on a wafer. In general, a number of wafers are processed at the same time, although some equipment designs operate on one wafer at a time (single-wafer processing).

The nomenclature of integrated circuits has changed as the complexity of ICs has increased: small-scale integration (SSI) has evolved to medium-scale integration (MSI), to large-scale integration (LSI), and to the mature very-large scale integration (VLSI), which has 10^5 or more devices per chip. The next generation of ICs are classified as ultra-large scale integration (ULSI).

A new generation of IC technology has developed roughly every three years. The Rule of Two holds that approximately for every two generations (six years), the device feature size decreases by two, and other properties such as logic gate speed, chip area, power dissipation, and maximum input/output (I/O) pins increase by two (1). As of this writing, IC technology has advanced to 0.5- μm circuit geometries for volume production, and is reaching into the deep submicrometer (0.35 μm and smaller) dimensions. Forecasts predict an ambitious scaling of device dimensions of 0.25 μm by 1998, and 0.18 μm by 2001 (2). Moreover, the number of chips per wafer is increasing, as well as the size of the chips and wafers. Silicon wafers are now fabricated in sizes up to 200 mm in diameter, and undergoing development of 300–400 mm sizes (3,4). The increase in the number of components per chip for dynamic random access memory (DRAM) ICs over time is shown in Figure 1 (5). The decrease in feature size and increase in chip area are shown in Figure 2a; the changes in cost and speed per device are shown in Figure 2b.

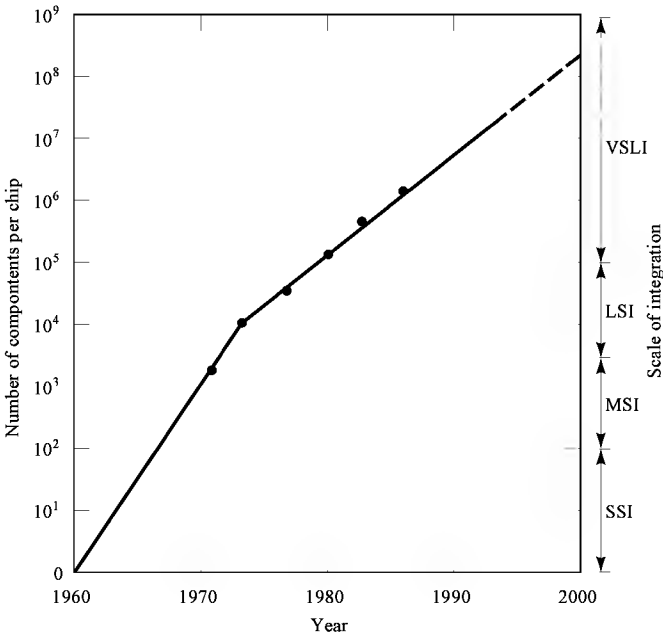


Fig. 1. Change in number of components per chip where () represents projected growth. See text.

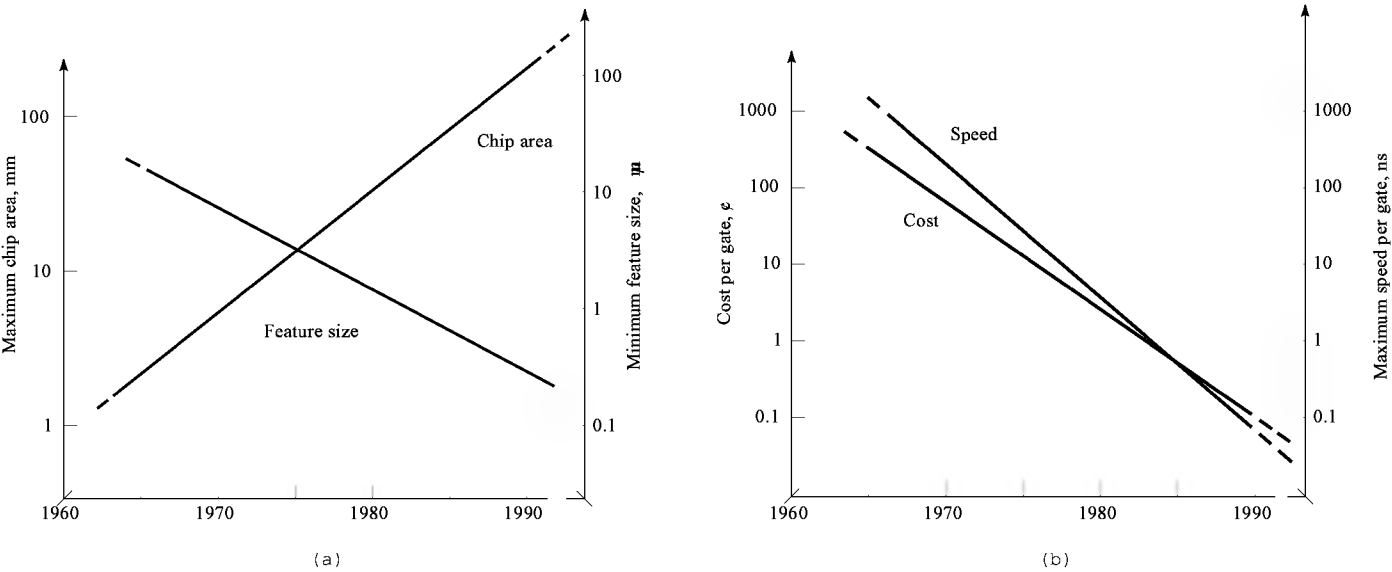


Fig. 2. Improvement over time: (a) chip area and minimum feature size; (b) maximum speed and cost per logic gate.

Courtesy of Custom VLSI Microelectronics.

Trends in the industry include transferring more of the functions that are found on the supporting printed circuit boards onto the wafers themselves, to reduce the amount of chip packaging and required interconnections (see ELECTRICAL CONNECTORS; PACKAGING, SEMICONDUCTORS AND ELECTRONIC MATERIALS). This development is known as wafer-scale integration (WSI) where the goal is to design a complete computer on a wafer (see COMPUTER TECHNOLOGY). These developments have been supported by concurrent advances in computer-aided design (CAD) tools, both in software and hardware, which have been used to develop integrated CAD design systems that perform IC layout design, simulation, and testing.

Silicon [7440-21-3], Si, technology predominates in the semiconductor industry (see SILICON AND SILICON ALLOYS). Gallium arsenide [1303-00-0], GaAs, is considered a possible substitute for silicon substrates, based on its potential for high speed applications where it can operate at high (1.9 GHz) frequencies using low power consumptions and high sensitivity (see ELECTRONIC MATERIALS). One reason that GaAs technology has not fulfilled its promise is that silicon technology has dramatically improved in the interim, particularly with improvements in speed, and has reduced the cost-effectiveness of pursuing GaAs development. Expense has limited usage of GaAs to microwave devices, primarily for military use (see MICROWAVE TECHNOLOGY). However, nonmilitary applications for GaAs devices have been growing, particularly in wireless products such as cellular phones (6).

Basics of Silicon Technology

VLSI technology is based on the unique attributes of silicon which have allowed the rapid evolution of integrated circuits. As a semiconductor source silicon has an adequate bandgap in its electronic make-up for the movement of electrons, it forms a stable insulating oxide, is abundant, and is inexpensive to make. In contrast, the use of GaAs has been limited to specialty applications because of its costly source and fragile nature (see GALLIUM AND GALLIUM COMPOUNDS).

The property that allows silicon to function in a number of capacities is electronic configuration. Silicon has four electrons in its outer shell. Its crystalline structure allows other elements to reside next to silicon and share electron orbitals with silicon, altering its electrical properties. These other elements are called dopants, and are introduced into the silicon structure through doping processes. The most common dopant is boron [7440-42-8], B, which has three electrons in its outer shell. Silicon doped with the electron-deficient boron has an overall positive charge, and is called *p*-type silicon. A second common dopant is phosphorus [7723-14-0], P, having five electrons in its outer shell. Silicon doped with P has an overall negative charge, and is called *n*-type silicon. Arsenic [7740-38-2], As, is another commonly used dopant.

The fabrication of an integrated circuit involves the sequential formation of alternating layers of insulators, semiconductors, and conductors on a silicon wafer. These layers are assembled to form transistor devices that are interconnected to produce particular electrical functions. The layers can be formed by deposition of new material, oxidation of material present on the surface, implantation of additional constituents into surface features, or epitaxial growth of silicon. In order to interconnect the layers, isolate devices from each other, and form integrated circuitry, these layers must be selectively patterned. The patterning is accomplished by photolithography and etching processes (see LITHOGRAPHIC RESISTS).

There are two kinds of integrated circuits (ICs): analogue, or linear ICs, and digital, or logic ICs. Analogue ICs produce, amplify, or respond to various voltages, and are used for any kinds of amplifiers, timers, oscillators, etc. Digital ICs respond to or produce signals that have only two voltage levels. These are used for microprocessors, memories, and microcomputers. It is possible to combine digital and analogue devices on one chip.

Digital IC families are further divided by design and function. The principal IC technologies include *p*- and *n*-channel metal-oxide semiconductors (PMOS and NMOS, respectively), complementary metal-oxide semiconductors (CMOS), bipolar, and integrated-injection-logic (I²L) devices. Of these, CMOS designs are by far the most popular, having an estimated 73% of the worldwide market in 1994 (7). Leading-edge microprocessors, application specific integrated circuits (ASICs), and DRAM ICs larger than 1 megabyte (Mb) are almost entirely fabricated with CMOS technology. Bipolar devices are the choice technology for high speed applications.

There are several reasons for the widespread use and development of CMOS devices, including low power density, relatively good noise immunity and soft error protection, design simplicity, and the capability to include lower power analogue and digital circuitry on the same chip. The most attractive feature has been the ability to scale CMOS technology to smaller dimensions. Processes exist that produce 0.5-μm dimensions under manufacturing conditions. Development is underway to effect 0.18-μm regimes (2). Another significant development in CMOS technology has been to reduce the power supply voltage from 5.5 to 3.3 V. There is expectation to step down further to 2.5 V as geometries decrease (1,8). Developments in equipment and processes are necessary to attain these goals.

Typical CMOS devices use both NMOS and PMOS transistors to form logic devices. A simple NMOS transistor is shown in Figure 3a. Two channels of *n*-doped silicon are formed in *p*-silicon to form a source and drain. An NMOS transistor is designed to permit a negative charge to move from the source to the drain in response to a positive charge in the gate. When the charge on the gate is large enough such that the source-to-gate voltage is higher than a threshold voltage *V_t*, electrons create a conducting path between drain and source causing current to flow. A common CMOS design combines NMOS and PMOS constructions in twin-well (twin-tub) structures, as shown in Figure 3b.

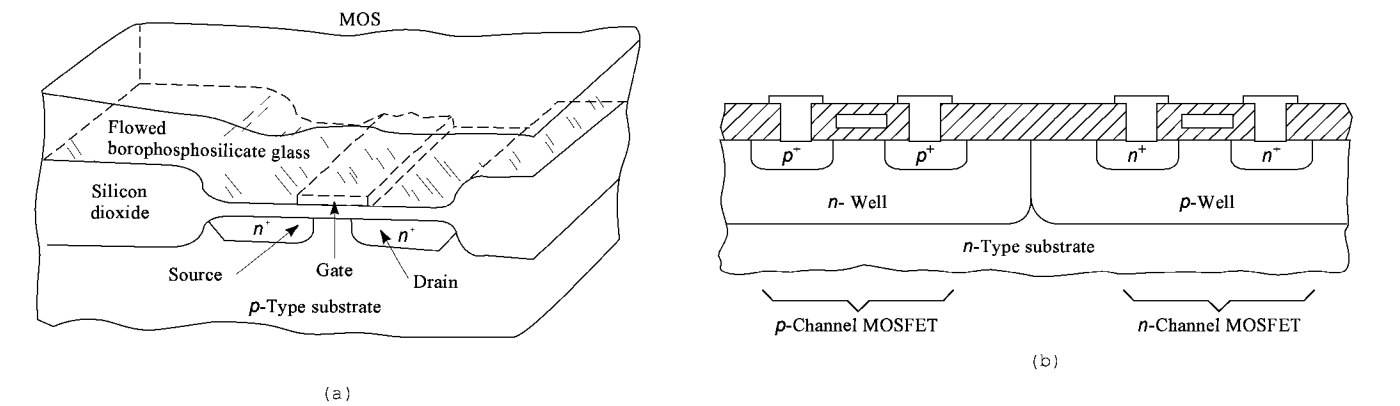


Fig. 3. Cross sections of electronics devices used in ICs. (a) NMOS transistor; (b) a twin-tub CMOS device on an *n*-type substrate.

Courtesy of Custom VLSI Microelectronics.

Crystal Growth and Wafer Preparation

The single-crystal silicon that is used in IC technology starts with a polycrystalline material called electronic-grade silicon (EGS). Its purity is determined by resistivity measurements made on the test ingot, or by low-ir absorption measurements. EGS is made by starting with a relatively pure form of sand, SiO₂. The sand is melted to form metallurgical-grade silicon (MGS), which reacts with HCl gas to form trichlorosilane [10025-78-2], SiHCl₃, gas. The SiHCl₃ reacts with H₂ in a chemical vapor deposition (CVD) process to form EGS.

$$2 \text{ SiHCl}_3 \text{ (g)} + 2 \text{ H}_2 \text{ (g)} \rightarrow 2 \text{ Si (s)} + 6 \text{ HCl (g)}$$

An alternative process involves pyrolysis of silane [7803-62-5].



This latter process is lower in cost and has fewer harmful by-products (9).

The polycrystalline EGS is converted to single-crystal silicon via the Czokralski (CZ) crystal growing process, based on the solidification of silicon atoms from the liquid phase at a moving interface. Volume production of 200-mm diameter crystals is standard. Development of crystals having diameters of up to 400 mm has been predicted (3).

The process of growing a pure crystal is sensitive to a host of process parameters that impact the incorporation of impurities in the crystal, the quality of the crystal structure, and the mechanical properties of the crystal rod. For example, the crystal-pulling mechanism controls the pull rate of the crystallization, which affects the incorporation of impurities in the crystal, and the crystal rotation, which affects the crystal structure.

Two common impurities that must be controlled in silicon crystal growth are oxygen and carbon. Oxygen affects the formation of donor regions, the yield strength of the crystal, and the level of defect generation. Oxygen, as silicon tetraoxide [12359-25-0], SiO₄, can act as a donor and change the resistivity of the silicon. Carbon also contributes to the formation of defects, and comes from graphite parts of the melt furnace.

Ingots of EGS are evaluated for resistivity, crystal perfection, and mechanical and physical properties, such as size and mass. The ingots are sliced into wafers using at least 10 machining and polishing procedures. These wafers are sliced sequentially from the ingot, and evaluated for the correct surface orientation, thickness, taper, and bow. As a final procedure, the wafers are chemically cleaned to remove surface contaminants prior to use.

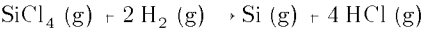
Defects in the silicon crystals affect the electrical, optical, and mechanical properties. Possible defects include point defects, which affect the kinetics of diffusion and oxidation; line defects, which can affect diffusion; planar defects; and volume defects, where impurities have precipitated and can act as sites for dislocation generation. Integrated circuits are typically constructed on wafers having <100> crystal orientation. Wafers 200-mm in diameter are being fabricated primarily for high volume, large-area circuits such as 16 Mb DRAMs, although 150 mm and smalled-sized wafers form the majority of product substrates. It is possible to fabricate almost twice as many 16 Mb chips on a 200-mm wafer as on a 150-mm wafer; the higher number of chips offsets the additional costs of the more expensive equipment and reduced throughput found in 200-mm processing (3). Important factors in the production of 200-mm wafers include a greater sensitivity to flatness, thermal stress, uniformity, and surface microroughness. Particle contamination control has become essential at the wafer fabrication stage to prevent damage in subsequent process steps. IC manufacturers are requiring very low particle counts from wafer suppliers, on the order of 50 or less particles per square centimeter, 0.1–0.15 μm in diameter (3).

Fabrication Processes

Deposition Processes.

Silicon Epitaxy. A critical step in IC fabrication is the epitaxial deposition of silicon on an integrated circuit. Epitaxy is defined as a process whereby a thin crystalline film is grown on a crystalline substrate. Silicon epitaxy is used in bipolar ICs to create a high resistivity layer on a low resistivity substrate. Most epitaxial depositions are done either by chemical vapor deposition (CVD) or by molecular beam epitaxy (MBE) (see T_{HIN FILMS}). CVD is the mainstream process.

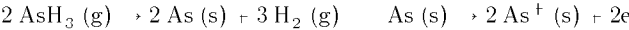
The CVD process consists of placing silicon wafers into a reactor chamber where process gases are introduced and heated to a high temperature. This induces a series of chemical reactions that result in the deposition of the desired epitaxial layer on the wafer substrates. Any of four gases may be used for silicon epitaxy: silicon tetrachloride [10026-04-7], SiCl₄, dichlorosilane [4109-96-0], SiH₂Cl₂, trichlorosilane, SiHCl₃, and silane, SiH₄. The most widely used and studied is silicon tetrachloride. The reaction that occurs is a hydrogen reduction.



This reaction is not a simple one. There are a number of intermediate chlorosilanes generated by competing reactions (10). The process is sensitive both to the thermodynamics and kinetics of the chemical reactions, and to the fluid mechanics (qv) of the gas flow in the reactor. The overall procedure involves purging the reactor with hydrogen gas, raising the temperature of the reactor, cleaning the wafers with a brief HCl etch, and replacing the HCl with the silicon source gas. A complete process cycle can take up to an hour.

CVD reactors can have one of several configurations. Each has particular advantages and disadvantages. Reactors that support wafers horizontally have difficulty controlling the deposition uniformity over all the exposed wafers. Reactors having vertical wafer support produce uniform deposition, but are mechanically complex. Barrel reactors are not suited for extended operation at temperatures greater than 1200°C.

It is often necessary to introduce dopant atoms into the epitaxial (epi) layers. Typically, the dopant sources are hydrides (qv) of the impurity atoms. Common dopants are boron hydride, ie, diborane(6) [19287-45-7], B₂H₆, for *p*-type doping, and arsine [7784-42-1], AsH₃, and phosphorus hydrides for *n*-type doping (11). For example:



Autodoping occurs when dopants are unintentionally released from a substrate through diffusion and evaporation, and subsequently reincorporated during the deposition layer. Epitaxial layers are typically doped at concentrations of 10¹⁴–10¹⁷ atoms cm^{−3}. The higher levels of doping are used in bipolar technology where the epi layer forms the transistor base. The epitaxial layer can be up to several hundred micrometers, and as thin as 0.05–0.5 μm. Uniformities of ±5% are common.

Molecular beam epitaxy is a non-CVD epitaxial process that deposits silicon through evaporation. MBE is becoming more common as commercial equipment becomes available. In essence, silicon is heated to moderate temperature by an electron beam in a high vacuum (10^{−6}–10^{−8} Pa (10^{−8}–10^{−10} Torr)) condition such that the volatile species travels at a relatively high velocity to the substrate wafer. The growth rate is 0.01–0.3 μ min^{−1} which starts to be competitive with CVD deposition rates.

The MBE process has the disadvantage of being expensive and having low throughput (the number of processed wafers per unit time). The principal advantage of MBE over CVD is that the former uses temperatures in the range of 400–800°C, reduces diffusion from substrate layers and consequent autodoping and this permits more accurate control of doping levels.

Oxidation of Silicon. Silicon dioxide [7631-86-9], SiO₂, is a basic component of IC fabrication. SiO₂ layers are commonly used as selective masks against the implantation or diffusion of dopants into silicon. SiO₂ is also used to isolate one device from another. It is a component of MOS devices, and provides electrical isolation of multilevel metallization structures (12). A comparison of Si and SiO₂ properties is shown in Table 1.

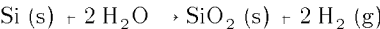
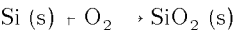
Table 1. Physical Properties of SiO₂ and Si

Property	SiO ₂	Si
density, g cm ^{−3}	2.27	2.33
dielectric constant	3.9	11.7
mp, °C	1700 ^a	1415
breakdown field, V μm ^{−1}	600 ^a	30 ^a
specific heat, J (gK) ^b	1.0	0.7
thermal conductivity, W (mK) ^{−1}	1.4	150
thermal diffusivity, cm ² s ^{−1}	0.006	0.9

linear coefficient of expansion, 10 ⁻⁶ K	0.5	2.5
etch rate, in buffered ^c HF at 27°C, μm min	0.10 ^a	0 ^a
band gap, eV	9	1.11
resistivity, Ω cm	>10 ¹⁶ ^d	10 ⁻³ to 10 ⁵ ^e

^a Approximate value.
^b To convert J to cal, divide by 4.184.
^c in 34.6% NH₄F, 6.8% HF, 58.6% H₂O buffer at 27°C
^d Insulator.
^e Semiconductor.

Silicon dioxide layers can be formed using any of several techniques, including thermal oxidation of silicon, wet anodization, CVD, or plasma oxidation. Thermal oxidation is the dominant procedure used in IC fabrication. The oxidation process selected depends on the thickness and properties of the desired oxide layer. Thin oxides are formed in dry oxygen, whereas thick (>0.5 μm) oxide layers are formed in a water vapor atmosphere (13).



The oxidation of surface silicon consumes the surface silicon, and the Si-SiO₂ interface moves down into the bulk silicon as the silicon is converted to the oxide. Oxygen atoms are transported from the gas phase through the oxide layer to the Si-SiO₂ interface where reaction with silicon takes place. Because oxygen atoms are incorporated into the silicon crystal structure, the volume of the resulting oxide is approximately 0.44 times the thickness of original silicon that is consumed. Thermal oxidation of silicon is not practicable when the surface silicon is very thin, as would be found in advanced IC designs. Such cases require the deposition of SiO₂ onto the silicon.

Typical oxide growth takes place at 101 kPa (1 atm) pressure in a horizontal diffusion tube apparatus, although vertical diffusion equipment is also used. Oxidation can be performed at higher pressures, such as 0.5–1 MPa (5–10 atm). The higher pressures permit the use of lower temperatures, which reduces migration of dopants in underlying layers. The growth rate is much higher under these conditions, and is used with advanced MOS applications. Growth rate is also dependent on the crystallographic orientation of the silicon surface because the orientation determines the concentration of the surface silicon atoms. Consequently, <111> silicon oxidizes faster than <100> silicon (14).

Silicon dioxide properties depend on the techniques used for oxide growth. The index of refraction for dry oxides decreases when higher processing temperatures are used whereas the oxide density increases.

Water as an impurity accelerates the oxidation rate. Figure 4 compares growth curves for SiO₂ under dry and steam conditions. Halogens can also be introduced to the oxidation process, thereby reducing sodium ion contamination. This improves dielectric breakdown strength, and reduces interface trap density (15).

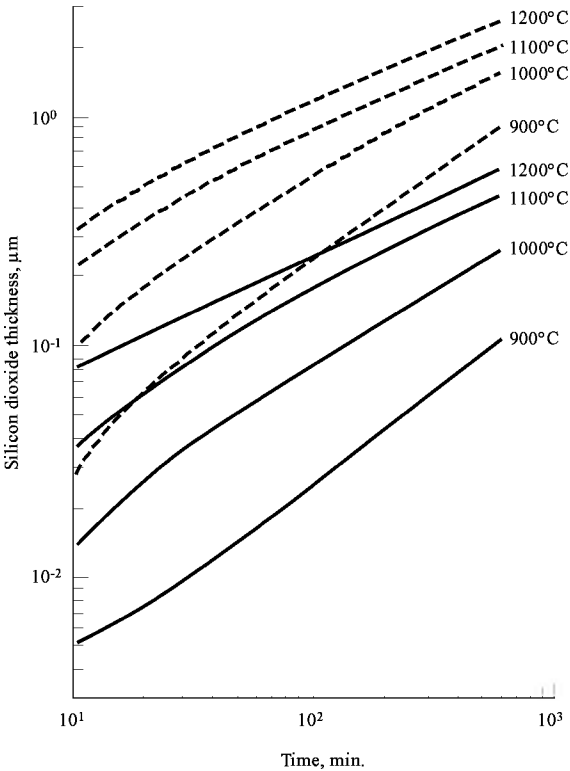


Fig. 4. Silicon dioxide growth rate using a <100> silicon substrate where the solid lines represent a dry oxygen and the dashed lines a steam atmosphere.

In the oxidation process, a layer of dopant is applied to the surface of silicon and patterned silicon dioxide for subsequent thermal diffusion into the silicon. The masking property of the SiO₂ is based on differences in rates of diffusion. Diffusion of dopant into the oxide is much slower than the diffusion into the silicon. Thus, the dopants reach only the silicon substrate. Oxide masks are usually 0.5–0.7 μm thick.

As the width and thickness of IC layers and patterns continue to shrink into the submicrometer range, SiO₂ layers need to be fabricated of 5–20 nm thickness. These thin oxides have properties that are very sensitive to the substrate cleanliness and uniformity, gas purity, and temperature control.

Dielectric Film Deposition. Dielectric films are found in all VLSI circuits to provide insulation between conducting layers, as diffusion and ion implantation (qv) masks, for diffusion from doped oxides, to cap doped films to prevent outdiffusion, and for passivating devices as a measure of protection against external contamination, moisture, and scratches. Properties that define the nature and function of dielectric films are the dielectric constant, the process temperature, and specific fabrication characteristics such as step coverage, gap-filling capabilities, density stress, contamination, thickness uniformity, deposition rate, and moisture resistance (2). Several processes are used to deposit dielectric films including atmospheric pressure CVD (APCVD), low pressure CVD (LPCVD), or plasma-enhanced CVD (PECVD) (see PLASMA TECHNOLOGY).

The most commonly used dielectric material is SiO₂, which may be deposited by several mechanisms, each distinguished by the specifics of the chemical source, the deposition process, and subsequent applications. Basic SiO₂ can be modified by the addition of boron, phosphorus, or both. SiO₂ and SiO₂-based films can be formed from thermal oxides; silane, SiH₄; tetraethoxysilane (TEOS), Si(OC₂H₅)₄; borophosphosilicate glass (BPSG); and spin-on glass (SOG) processes. The majority of CVD SiO₂ films are made from SiH₄ or TEOS.

Gate oxide dielectrics are a crucial element in the down-scaling of *n*- and *p*-channel metal-oxide semiconductor field-effect transistors (MOSFETs) in CMOS technology. Ultrathin dielectric films are required, and the 12.0-nm thick layers are expected to shrink to 6.0 nm by the year 2000 (2). Gate dielectrics have been made by growing thermal oxides, whereas development has turned to the use of oxide nitride oxide (ONO) sandwich structures, or to oxynitrides, SiO_{*x*}N_{*y*}. Oxynitrides are formed by growing thermal oxides in the presence of a nitrogen source such as ammonia or nitrous oxide, N₂O. Oxidation and nitridation are also performed in rapid thermal processors (RTP), which reduce the temperature exposure of a substrate.

Another important use of dielectrics is as intermetal dielectrics (IMDs), where the dielectrics insulate metal lines from each other. The dielectric material must fill small gaps with high aspect ratios (depth to width) while maintaining all other dielectric properties. It is essential that the IMDs are void-free at submicrometer dimensions for both performance and reliability.

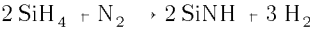
Historically, SOG techniques have been used the most for IMD fabrication, but TEOS ozone (TEOS O₃) processes are more recent developments that have been increasing in popularity based on excellent step coverage and void-free characteristics. TEOS O₃ doped with boron and phosphorus (BPTEOS O₃) has replaced BPSG in small-scale devices, and has been used successfully in 4- and 16-Mb DRAM production (16).

Dopant species can be codeposited with the SiO₂ by introducing small amounts of the dopants in hydride or halide form. P-doped SiO₂, called P-glass, functions as an insulator between polysilicon gates and the top metallization layer of ICs. It is also used as a final passivation layer over devices, and as a gettering source (17).

There are two types of deposited films known as silicon nitride. One is deposited via plasma-enhanced CVD at temperatures <350°C (18). In this process silane and ammonia react in an argon plasma to form silicon inide [14515-04-9], SiNH.



Alternatively, silane can react in a nitrogen discharge.



Plasma-deposited silicon nitride contains large amounts of hydrogen, typically in the range of 20–25 atomic % H, and has polymer-like properties. The electrical resistivity of the film depends on the deposition temperature, the film stoichiometry, and the amounts of hydrogen and oxygen in the film.

A second type of silicon nitride, called stoichiometric silicon nitride, is deposited at much higher temperatures using CVD or LPCVD in the form of Si₃N₄. Stoichiometric silicon nitride can be used as a mask for the selective oxidation of silicon. Here the silicon nitride is patterned over a silicon substrate, and the exposed silicon is oxidized. The silicon nitride oxidizes very slowly compared to the silicon.

Polysilicon. Polysilicon is used as the gate electrode material in MOS devices, as a conducting material for multilevel metallization, and as contact material for devices having shallow junctions. It is prepared by pyrolyzing silane, SiH₄, at 575–650°C in a low pressure reactor. The temperature of the process affects the properties of the final film. Higher process temperatures increase the deposition rate, but degrade the uniformity of the layer. Lower temperatures may improve the uniformity, but reduce the throughput to an impractical level.

The structure of the polysilicon depends on the dopants, impurities, deposition temperature, and post-deposition heat annealing. Deposition at less than 575°C produces an amorphous structure; deposition higher than 625°C results in a polycrystalline, columnar structure. Heating after deposition induces crystallization and grain growth. Deposition between 600 and 650°C yields a columnar structure having reasonable grain size and <110>-preferred orientation.

When required, arsenic, phosphorus, or boron dopants can be added in subsequent steps by diffusion or ion implantation (qv), or by adding dopant gases during deposition. All three methods are used in IC fabrication. The doping elements reduce the resistivity of the polysilicon. The degree of change of resistivities is a function of the deposition temperature, the dopant concentration, and annealing temperature. The dopant concentration affects the etch rate in plasma etching. Oxygen can be used to dope polysilicon to increase its film resistivity, creating a semi-insulating polysilicon (SIPOS) that can be used as a passivating coating for high voltage devices (see ELECTRONICS, COATINGS). SIPOS is formed by depositing silane and nitrous oxide [10102-43-9] together at 600–700°C. A metal or metal silicide, eg, tungsten or tungsten silicide [12039-88-2], may be deposited over the polysilicon to increase its conductivity.

Polysilicon is also oxidized to SiO₂ in dry oxygen at 900–1000°C to form an insulator between the doped polysilicon gate and other conducting layers. Because the surface of the polysilicon is rough relative to the surface of a single-crystal silicon, the SiO₂ grown on polysilicon is itself rough, resulting in lower breakdown fields, higher leakage currents, and higher stresses.

Metallization. Integrated circuits require conductive layers to form electrical connections between contacts on a device, between devices on a chip, between metal layers on a chip, and between chips and higher levels of interconnections needed for packaging the chips. It is critical to the success of IC fabrication that the metallization be stable throughout the process sequence in order to maintain the correct physical and electrical properties of the circuit. It must also be possible to pattern the blanket deposition.

In metal-oxide semiconductor field-effect transistor (MOSFET) devices the control of the on/off function occurs at the gate electrode, which is usually polysilicon. The gate is isolated from the substrate silicon by an insulating layer such as thermally grown SiO₂. The two regions adjacent to the gate are the source and the drain. These features are usually interconnected to first-level aluminum [7429-90-5] metallization through contacts, where the metallization controls the speed of the circuit, and determines the gate-to-source voltage that switches the MOSFET on or off. Vias are small holes in the circuitry that interconnect the first-level metal to the second-level metal, the second to the third, and so on.

Metallization layers are generally deposited either by CVD or by physical vapor deposition methods such as evaporation (qv) or sputtering. In recent years sputter deposition has become the predominant technique for aluminum metallization. Energetic ions are used to bombard a target such as solid aluminum to release atoms that subsequently condense on the desired substrate surface. The quality of the deposited layers depends on the cleanliness and efficiency of the vacuum systems used in the process. The mass deposited per unit area can be calculated using the cosine law of deposition:

$$R_D = M_e \pi r^2 \cos \phi \cos \theta$$

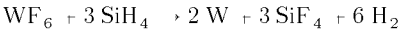
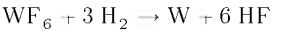
where *R_D* is the evaporation rate, *M_e* is the total mass of the evaporated material, and *r*, *φ*, and *θ* relate the angles between the source and the surface of deposition (19).

A large number of factors can degrade the desired properties of the metallization layer. Impurities in the film, poor adhesion to the substrate, grain size, step coverage, and thickness nonuniformity all reduce the effectiveness of the deposition. Stresses in the layer can occur from differences in TCE between the metal layer and the substrate or from intrinsic stress. Both can result in crack formation. Step coverage can be improved by heating the substrate, optimizing the orientation between the substrate and the source, and using larger-area sources (20).

Aluminum, the most common material used for contacts, is easy to use, has low resistivity, and reduces surface SiO₂ to form interfacial metal-oxide bonds that promote adhesion to the substrate. However, as designs reach submicrometer dimensions, aluminum, Al, has been found to be a poor choice for metallization of contacts and via holes. Al has relatively poor step coverage, which is nonuniform layer thickness when deposited over right-angled geometric features. This leads to keyhole void formation when spaces between features are smaller than 0.7 μm. New collimated sputtering techniques can extend the lower limit of Al use to 0.5-μm applications.

Tungsten CVD (WCVD) is the technique of choice to fill contact and via holes in sub-0.5 μm geometries. Tungsten has extremely good step coverage, good resistance to electromigration, low resistivity, and good adhesion to underlying silicon (in the case of contacts) and aluminum (in the case of

vias) (21). The chemistries that are used most frequently in tungsten deposition are mixtures of silane or hydrogen with tungsten hexafluoride, WF₆, where the silane and hydrogen act as reducing agents for the tungsten source. The overall equations are (22):



The blanket deposition is then sputter etched through a resist to pattern the metallization. Selective deposition of W, under development, would deposit metal only in desired areas, and would reduce process steps and costs.

Copper is an attractive metallization element because of its high conductivity. It has been added to Al in low concentrations (AlSi(1^o)-Cu(0.5^o)) to improve conductive priorities. Selective, low temperature copper CVD processing, using copper(I) β-diketonate compounds, has been carried out (23).

Impurities that can negatively affect the physical and electrical properties of the metallization layer can originate from several sources, particularly the deposition source and the gaseous environment. Impurities stemming from the source bombard the surface of the growing film and get trapped in the metal layer.

Finally, the metallization layer usually requires patterning, which can be done by reactive ion etching (RIE) or back-sputtering. The two processes are similar. In both techniques accelerated ions hit the substrate and forcibly detach atoms or molecules from the surface. RIE uses reactive gases such as chlorine, Cl₂ or trichlorofluoromethane [75-69-4], CCl₃F. Inert gases such as argon or neon are used in back-sputtering.

Doping of Layers.

Diffusion. Another technique for modifying the electrical properties of silicon and silicon-based films involves introducing small amounts of elements having differing electrical compositions, dopants, into substrate layers. Diffusion is commonly used. There are three ways dopants can be diffused into a substrate film: (1) the surface can be exposed to a chemical vapor of the dopant at high temperatures, or (2) a doped-oxide, or (3) an ion-implanted layer can be used. Ion implantation is increasingly becoming the method of choice as the miniaturization of ICs advances. However, diffusion is used in VLSI technology (24,25).

Theoretical studies of diffusion aim to predict the distribution profile of an exposed substrate given the known process parameters of concentration, temperature, crystal orientation, dopant properties, etc. On an atomic level, diffusion of a dopant in a silicon crystal is caused by the movement of the introduced element that is allowed by the available vacancies or defects in the crystal. Both host atoms and impurity atoms can enter vacancies. Movement of a host atom from one lattice site to a vacancy is called self-diffusion. The same movement by a dopant is called impurity diffusion. If an atom does not form a covalent bond with silicon, the atom can occupy in interstitial site and then subsequently displace a lattice-site atom. This latter movement is believed to be the dominant mechanism for diffusion of the common dopant atoms, P, B, As, and Sb (26).

When the concentration of the solute is low, the measured diffusion profiles are predictable from Fick's second law of diffusion:

$$\frac{\delta C(x, t)}{\delta t} = D \frac{\delta^2 C(x, t)}{\delta x^2}$$

where *D* is the diffusivity, *C* is the concentration of solute, *x* is the coordinate axis in the direction of solute flow, and *t* is the diffusion time. Diffusivity at low concentration is called intrinsic diffusivity, *D_i*. When the concentration of the solute is > *N_i*, where *N_i* is the intrinsic carrier concentration, and *T* is temperature in Kelvin, Fick's law as given is insufficient to accurately describe the resulting dopant profile. Additional concentration-dependent diffusivities must be added to the equation. The diffusivities can be determined by experimentally obtained dopant profiles. At high solute concentrations, the silicon is considered as extrinsic silicon, and the diffusivity is called the extrinsic diffusivity *D_e*. When arsenic is diffused into silicon at concentrations below 10¹⁶ cm⁻² and at temperatures >1000°C, nearly all the arsenic ions are ionized and are electrically active. At higher concentrations, and lower diffusion temperatures, the ionized arsenic is only a fraction of the total arsenic. A similar effect is seen for boron (27).

Diffusivities of various elements are determined experimentally. Dopant profiles can be determined. The junction depth can be measured by chemically staining an angle-lapped sample with an HF-HNO₃ mixture. The *p*-type region of the junction stains darker than the *n*-type region. The sheet resistivity can also be measured using a four-point probe measurement. These two techniques are used for process monitoring.

More accurate dopant profile information can be obtained using additional methods. The capacitance-voltage (C-V) method is used to measure the reverse-bias capacitance as a function of the applied voltage. It is common to use Rutherford back-scattering (RBS) to determine the distributions of arsenic, platinum, or gold in silicon. This method cannot, however, be used for measuring boron or phosphorus profiles. Another surface-analytical technique used for these measurements is secondary ion mass spectrometry (sims), which has high sensitivity for many elements including B and As (see MASS SPECTROMETRY).

Ion Implantation. Ion implantation (qv) is a technique designed to introduce dopant species into silicon in a controlled, reproducible process. The dopant species are ionized and accelerated toward a substrate surface. The ions have sufficient energy to penetrate the surface and enter the crystal lattice. A series of collisions take place in a cascading effect that affects not just the dopant but also the atoms in the lattice. The depth of penetration, ie, the profile, is controlled by the accelerating energy of the ions. Accelerating energies can range from 1 keV to 1 MeV, allowing penetrations of 100 nm to 10 μm (28). Doping profiles for a bipolar transistor are shown in Figure 5.

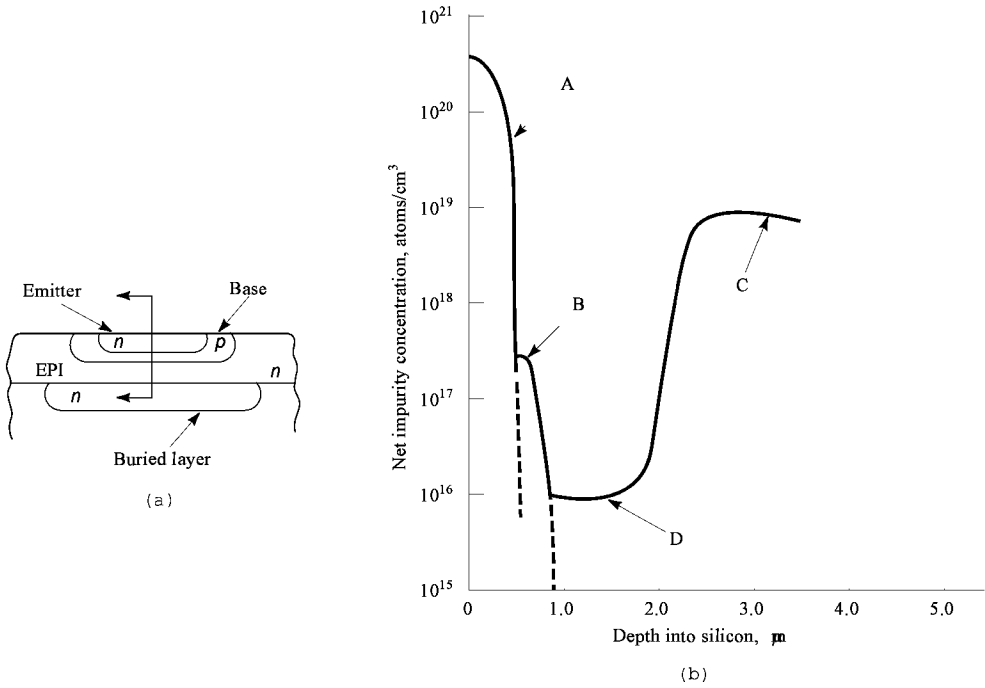


Fig. 5. Bipolar transistor (a) schematic and (b) doping profiles of A, arsenic ion implanted into the silicon of the emitter (*n*-type); B, boron ion implanted into the silicon of the base (*p*-type); C, antimony ion implanted into the buried layer (*n*-type); and D, the epi layer (*n*-type).

Typically, a source gas such as boron trifluoride [7637-07-2], BF₃, is exposed to an ion source that causes the gas to ionize. An analyzer discriminates between all the ionic particles using a magnetic field that can select particles having the correct mass-to-charge ratio to pass through the analyzer to an acceleration tube. The ions are accelerated in the tube and collimated into a beam that is scanned over the substrate wafer. The three primary parameters of any implantation process are the type of dopant species, the accelerating energy used for implantation, and the dose of the source gas. The dose is the total number of ions that enter the wafer. Dose, ϕ , can be calculated

$$\phi = \frac{1}{QA} \int I dt$$

where Q is the charge on the ionized species; A , the area of the target; and I , the acceleration beams, integrated over time t (29). Doses range from 10¹² ions cm² for threshold adjustment to 10¹⁸ ions cm² for buried insulators.

Upon entering a substrate crystal, the implanted ion initiates a series of elastic and inelastic collisions between atomic nuclei and atomic electrons, respectively. The initial collision is between the dopant ion and a silicon lattice atom. The accelerated dopant has sufficient energy to displace the lattice atom, which in turn can displace another lattice atom. An ion entering the wafer and having an initial energy of 100 keV has sufficient energy to initiate a cascade of displacements, because the binding energy of a lattice site is only 10–20 keV.

Channeling is a phenomenon of implantation where impacting ions enter into ordered open spaces, called channels, in the crystal lattice. Channels allow the implanted ion to avoid nuclear collisions and travel further into the lattice than would an ion that collided with atoms in the lattice. Channels can be planar or axial. The primary energy loss in channeled ions is from electron scattering. Channeling can affect the degree of control of the doping profile, especially when dealing with shallow junctions. Consequently, wafers are tilted to avoid channeling, typically at a 7° angle (30).

The primary sources of contamination in ion implantation come from metal atoms that may be etched off reactor fixtures, such as reactor walls, wafer holder, clips, and so on. The pump oils used by the vacuum pumps may be a source of hydrocarbon contamination. The dopant sources themselves are not a significant source of contamination because unwanted ions are separated out from the beam during beam analysis.

The implantation of dopant ions into the wafer crystal creates a significant disruption of the lattice structure. An annealing step is performed after implantation to repair the lattice and place the dopant ions into substitutional sites where the dopants can be electrically active with neighboring atoms. Not all of the implanted ions become active, and one primary measure of success of implantation is defined as the fraction of dopant that is electrically active. This is measured experimentally using a Hall effect technique which determines the average effective doping, ie, the integral over local doping densities and local mobilities evaluated per unit surface area

$$N_{\text{Hall}} = \frac{\left(\int_0^{x_j} \mu n dx\right)^2}{\int_0^{x_j} \mu^2 n dx}$$

where n is the mobility; n , the number of carriers; and x_j is the junction depth. If annealing activates all of the implanted atoms, N_{Hall} is then equal to the dose, ϕ (31).

The annealing must not only repair lattice structure, it must also do so under conditions that minimize diffusion. Furnace annealing is one method that has been used to anneal implanted wafers. In furnace thermal annealing, amorphous silicon regrows by solid-phase epitaxy. An interface exists between the amorphous and crystalline phases of the silicon, which rises to the surface during the annealing process. The velocity of the movement of the interface depends on the temperature, degree and type of doping, and the crystal orientation. Some impurities such as oxygen, carbon, nitrogen, or argon reduce the recrystallization rate. Boron, phosphorus, and arsenic increase the rate of annealing. Rapid thermal annealing (RTA) is a more recent technology developed for annealing purposes. RTA involves heating the wafer for very short exposures, from 100 down to nanosecond periods. The critical feature of RTA is to minimize diffusion in shallow junctions.

Two newer areas of implantation have been receiving attention and development. Focused ion beams have been investigated to allow very fine control of implantation dimensions. The beams are focused to spot sizes down to 10 nm, and are used to create single lines of ion-implanted patterns without needing to create or use a mask. Although this method has many attractive features, it is hampered by the fact that the patterning is sequential rather than simultaneous, and only one wafer rather than many can be processed at any one time. This limits the production applications of the technique.

Wafer charging is becoming a critical issue as new MOS devices are designed with thinner (<10 nm) gate dielectrics. Lower energy implanters require optics that compensate for beam divergence which occurs at low energies (32).

Devices are being designed that have shallow junctions, where the dopants are introduced less than 100 nm beneath the surface. Shallow junctions are formed by using low energies for implantation. A technique being used for submicrometer implantation is large-angle tilt implanted drain (LATID) and source. High (200 keV to 3 MeV) energy implanters are also under development for profile and defect engineering. This implant process is used to produce buried oxide layers, thereby reducing costs by eliminating the need to form separate epi layers (33).

Patterning.

Lithography. Integrated circuits require that the conducting and nonconducting layers are patterned to form the actual circuits and interconnections needed for a functional device. Typically, the patterns in the functional layers are introduced by first depositing a blanket layer of a polymeric material, called a resist, over the wafer surface. A discrete layer containing the desired pattern information, called the mask, is placed over the resist. The resist is irradiated through the mask with one of several possible media, which include visible or ultraviolet light, electrons, x-rays, or ions. The resist responds to the irradiation by either polymerizing in the exposed regions (negative resists), or preventing polymerization where exposed (positive resist). The unpolymerized portion of the resist is removed in a developer solution. A pattern now exists on the surface of the wafer that can be transferred to the underlying, exposed functional layer by etching the functional layer in the areas not covered by the resist. Alternatively, an additional layer can be selectively deposited into the areas not covered by the resist (see LITHOGRAPHIC RESISTS).

Misregistration between one circuit layer and another is critical for high yield and reliability in IC fabrication. Misregistration can occur at many points during fabrication; for example, when two masks do not align perfectly, or the printer that forms the mask adds to the misalignment. Processing the wafer adds a multitude of possible sources for misregistration, such as etch processing or wafer distortion. The ratio of minimum feature size to registration tolerance is usually in the range of 3–5, permitting high precision of micrometer or submicrometer features. If the contributions are independent, then the total registration error, Δ_r is given by

$$\Delta_r = \left(\sum_i \Delta_i^2\right)^{1/2}$$

where Δ_i is an individual error.

Optical Lithography. Optical lithography uses visible or ultraviolet light as the exposure media, and is the dominant lithographic process used for patterning IC wafers. The linewidth limit is near 0.4 μm , although some narrower features may be possible (34). The masks typically are made from patterned, opaque chromium films on glass.

Two types of resists are used: negative and positive resists. Negative resists have a chemically inert polymer component, such as rubber, and a photoreactive agent that reacts with light to form cross-links in the rubber. When placed in an organic developer solvent, the unexposed, unpolymerized resist dissolves, leaving a polymeric pattern in the exposed regions. Because the polymer swells in the solvent, the resolution is limited to two to three times

the film thickness. Swelling is a critically limiting factor in all negative resists, limiting line resolution. The swelling between two closely spaced lines of resists can create a bridge between the lines, and weaken the adhesion of the resist line to the substrate. An example of a patterning process using a negative resist is shown in Figure 6.

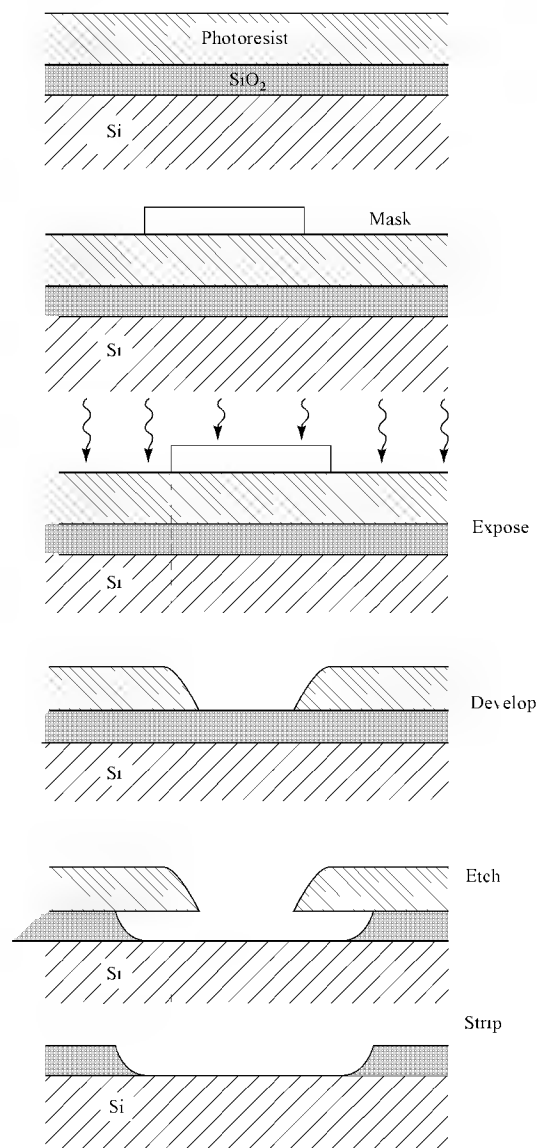


Fig. 6. Schematic illustration of the photolithographic patterning process used for defining features in silicon dioxide using (▨) a positive photoresist that polymerizes light, where (□) represents the mask; (■) SiO₂; and (▧) Si. Development includes removal of the mask and undeveloped photoresist.

Positive resists have as the photoreactive component a dissolution inhibitor that is destroyed in the regions exposed to the light. The resist is developed in an aqueous solution, where the exposed region dissolves away. The resists do not swell as much in the aqueous developer, allowing higher resolution.

Masks can be placed directly on the resist surface for contact printing allowing submicrometer line resolution. But this procedure may have distortions develop if there are any surface irregularities in the mask or substrate. Contact printing can also cause defects in the mask itself from the physical contact between the mask and the wafer. A second printing process, called projection printing, holds the mask away from the wafer. Projection printers can process 13–50 wafers per hour, depending on the size of the wafer. The different printing methods are shown in Figure 7.

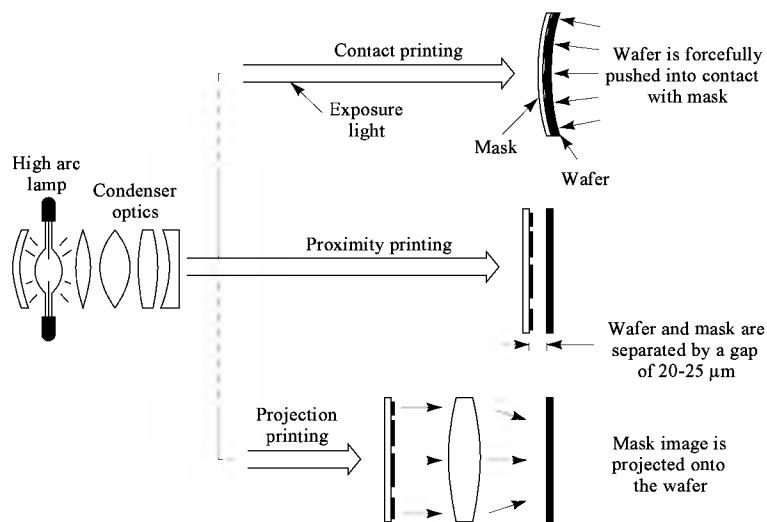


Fig. 7. Schematic illustration of different photolithographic exposure techniques.

The use of optical lithography for submicrometer resolutions has been extended by developments such as phase-shift lithography (35). A

phase-shift reticle (mask) is one which has a patterned layer added to the conventional chrome-on-glass lithographic reticle. Light passing through the glass layer emerges with a different phase than light passing through the phase layer, resulting in an interference pattern that heightens resolution and depth of focus. The next generation of optical lithography is projected to use deep-uv (248 nm) irradiation to achieve 0.20–0.15 μm resolution (32), which may be a necessary but very expensive transition. Manufacturable deep-uv resist technology requires further development (36).

Electron Lithography. It is possible to achieve better resolution using electron rather than optical lithography because of the small wavelength of 10–50 keV electrons. Although e-beam direct write systems have always shown great potential for producing fine patterns, production applications have been limited by low throughput to low volume applications such as mask generation (37).

Electron lithography uses both positive and negative resists as does optical lithography. The electron beam induces cross-links between molecules in negative resists, making the exposed areas less soluble in developer. In positive resists, the electron beam causes bond scission, resulting in lower solubility of the resist in the exposed areas when placed in developer. Polymethyl methacrylate (PMMA) offers the highest resolution, and is used both in optical and electron lithography.

Scanning electron beam systems are available commercially, and are commonly used for mask generation. Electron projection systems are also used to obtain resolution over a large field. Current cathode sources have a short lifetime, limiting use in production processes.

X-Ray Lithography. The resists used in x-ray lithography are the same as those used in electron lithography, because an x-ray resist is exposed predominantly by the photoelectrons produced during x-ray absorption. The photoelectrons are of much lower (0.3–3 keV) energies than those used in electron lithography (10–50 keV) permitting much higher resolution of the mask image. A 1-μm thick hydrocarbon resist having density of 1 g cm⁻³ absorbs only about 10⁰ % of the x-ray flux at the Al_{Kα} wavelength 0.83 nm. This small absorption yields uniform exposure through the resist thickness. Gold is the most commonly used material for the mask, and is patterned by electroplating (qv) or ion milling (38). However, x-ray lithography has yet to be developed for production applications. It is not considered as cost-effective as deep-uv or phase-shift lithography. X-ray lithography is projected to see broader application toward the end of the 1990s as 1 Gb DRAM technology is predicted to require 0.2–18 μm resolutions (37).

Etching. After a resist is patterned on a wafer, the exposed or unwanted substrate is removed by etching processes. Subsequently the resist is removed, leaving a desired pattern in a functional layer of the integrated circuit. Etching is performed to pattern a number of materials in the IC fabrication process, including blanket polysilicon, metal layers, and oxide and nitride layers. The etch process for each material is different, and adapted to the material requirements of the substrate.

The goal of etching processes is to achieve precisely patterned geometries having vertical profiles by removing unwanted material from a substrate or blanket layer inflicting minimal damage to the remaining substrate. The desired etch characteristics are high etch rate, anisotropy (vertical profiles), uniformity, and high selectivity for the material that is to be removed over the material that acts as a resist. An example of isotropic etching is depicted schematically in Figure 8a, showing the undercutting that results from etching the exposed surfaces in all directions. The desired anisotropic etch is sketched in Figure 8b, showing the vertical profiles of the sidewalls after etching occurs only in the vertical direction.

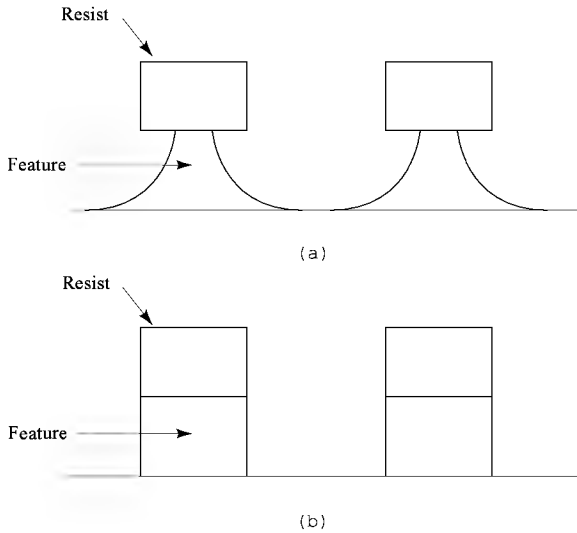
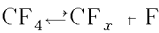


Fig. 8. Schematic of etching directionality showing (a) isotropic etch, and (b) anisotropic etch.

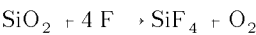
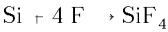
Etching is performed either by plasma or reactive ions. In plasma etching, a plasma is generated by applying a d-c potential across two conducting electrodes in the presence of a neutral gas. The gas becomes ionized into electrons, singly or multiply charged ions, and neutral atoms, molecules, and molecular fragments. The plasma species physically or chemically interact with the solid surfaces forming volatile products that are removed from the wafer surface.

The physical removal of surface material is called sputtering, where energetic, chemically inert ions such as Ar⁺ or Xe⁺ are accelerated toward the wafer and physically eject material from the surface. The yield is defined as the ratio of the number of ejected surface atoms to the number of incoming ions per given ion energy.

Chemical removal of surface material is produced through standard bond-breaking reactions. Typically chlorofluorocarbons (CFCs) have been used, eg, CFCl₃, CF₂Cl₂, CF₃Cl, CF₄, CHF₃, C₂ClF₅. For example, CF₄ dissociates into F atoms and fluorinated fragments of CF_x in a plasma:



where $x \leq 3$. Oxygen is added to the plasma mix to react with the CF_x molecules and drive the reaction to the right. In the case of etching Si or SiO₂, F is the active etchant species, reacting with Si and SiO₂ to form volatile SiF₄ (39).



A critical feature of plasma etching is managing the directionality of the etching reaction, relative to the resist and the surface layer. Isotropic etching occurs when the etching reactions are equal in all directions on the surface producing an effect called undercutting, as shown in Figure 8a. Undercutting is not a problem in patterns having feature height, where the degree of undercutting can be compensated for by making the spacing between the resist narrower to obtain the desired final spacing in the functional layer. Anisotropic etch occurs when the attack of the etching molecules is faster in the vertical direction than in the horizontal direction. This results in deeper channels relative to the width, and straighter wall profiles (Fig. 8b). This feature is of primary importance in VLSI fabrication, where narrow spacing is required.

Reactive ion etching (RIE) is an etch process whereby the wafers are held by the radio frequency-driven electrode instead of the grounded electrode, which becomes the chamber itself. The larger grounded area, in combination with lower operating pressures, leads to significantly higher energy

bombardment. An advantage of this system is that wafer-to-wafer etching nonuniformities are reduced.

Changes in technology of plasma etching have been driven by two principal forces. The first is the scaling to submicrometer geometries, which has impacted on process design, gas chemistries, and plasma sources. In conventional plasma chemistries, the anisotropy of vertical etching was achieved by adding halocarbons such as CCl₄, CF₄, CHCl₃, or CHF₃ to the primary enchant gas to form passivating layers that would protect the sidewalls from etching. The typically 0.1-μm thick passivating layers are a significant liability when feature size is less than one micrometer and impossible in under 0.5-μm designs. Engineers having submicrometer etch requirements are turning to single-wafer etchers with precise temperature controls that circumvent the need for passivating layers (40).

Submicrometer etching requires gas chemistries having much greater selectivity and high etch rates. One solution is to use multistep processing, where a number of etch steps are used to etch through one layer. It is possible to have 10–15 procedures to etch through two to three layers. In polysilicon etching, the etch gases must have sufficient selectivity to etch stop over 20-nm thick (or thinner) oxide layers. Polysilicon etch gases have a selectivity to oxide in the range of 40:1 to 100:1 (40,41). Although bromine-based gases can provide the required selectivity, corrosiveness has prevented widespread use of these gases.

Plasma sources are also being introduced to produce plasmas at lower pressures and process temperatures. Inductively coupled plasma (ICP) and transformer-coupled plasma (TCP) are among the more commonly used sources, operating below 2.6 Pa (20 mTorr) (42). Low temperature RIE processing operates between 26–67 Pa (200–500 mTorr).

A second force that is seriously affecting etch technology is the move to actively eliminate CFC gases to comply with the Montreal Protocol for protection of the ozone layer. These gases, eg, CFCl₃, CCl₂F₂, CF₃Cl, and CF₃Br, are being replaced by fluorine-based gases for tungsten etch, such as SF₆, NF₃, and SiF₄; chlorine-based gases for Al etch, eg, Cl₂, BCl₃, and SiCl₄; and bromine-based gases, Br₂, and HBr. Table 2 lists conventional and newer chemistries.

Table 2. Trends in Plasma Etch Chemistry^a

Material being etched	Conventional chemistry		Replacement chemistry
polysilicon	Cl ₂ or BCl ₃ CCl ₄ ^b	SiCl ₄ /Cl ₂	no carbon contamination
	CF ₄ ^b	BCl ₃ Cl ₂	
	CHCl ₃ ^b	HBr Cl ₂ O ₂	increased se-lectivity to SiO ₂ and resist
	CHF ₃ ^b	HBr O ₂	no carbon contamination
		Br ₂ SF ₆	
		SF	higher etch rate
Al		CF ₄	
	Cl ₂ ^c	SiCl ₄ Cl ₂	better profile control
	BCl ₃ ^c	BCl ₃ Cl ₂	no carbon contamination
Al–Si(1 ^o o)–Cu (0.5 ^o o)	SiCl ₄ ^c	HBr Cl ₂	
	same as Al	BCl ₃ Cl + N ₂	N ₂ accelerates Cu etch rate
	SF ₆ Cl ₂ CCl ₄	SF only	no carbon contamination
W		NF Cl ₂	etch stop over TiW and TiN
		CF ₃ BR	higher selectivity trench etch
single-crystal Si	Cl ₂ or BCl ₃ ^c	HBr NF ₃	
		CCl ₂ F	all are CFC alternatives
	CCl ₂ F ₂	CHF ₃ CF ₄	
	CF ₄	CHF ₃ O ₂	
SiO ₂ (BPSG)	C ₂ F	CH ₃ CHF ₂	
	C ₃ F ₈	CF ₄ O ₂	
		CF ₄ H ₂	
		CHF ₃	
		CH ₃ CHF ₂	
Si ₃ N ₄	CCl ₂ F ₂		all are CFC alternatives
	CHF ₃		

^a Ref. 45.
^b Sidewall passivating gas.
^c Plus sidewall passivating gases.

Plasma etching can create a variety of damaging effects to a substrate. Typical damage effects include gate oxide breakdown, high reverse leakage current, low minority carrier lifetime, contamination, damage to the silicon surface charge, and lattice damage (43). The sources of the damage can be typically attributed to either or both of two effects: current flow-induced damage and plasma exposure damage. The first affects primarily the dielectric layers, where voltage across the dielectric produces wear-out from bond-breaking, trapping, or both. The second is a side effect of particle or photon flux impingement on the substrate materials (44). The regions that are most vulnerable to physical damage include the area in the oxide layer, at the layer interfaces, and in the silicon substrate. As of 1994, damage detection, measurement, and analysis are not adequate for submicrometer designs.

Process Integration. The fabrication of a VLSI integrated circuit involves sequential processing steps. Step interrelationships must be considered in designing a process sequence. The registration of one layer to another layer already present is repeated frequently with resist patterning followed by etching or implantation. Process temperature must be compatible with the condition of the substrate, such that a high temperature process does not damage an underlying layer or destroy a precision dopant profile.

Figure 9 shows a simplified fabrication sequence for an oxide-isolated *p*-well CMOS process that illustrates many of the essential steps used in IC manufacture. These steps are as follows:

- Step 1. Starting with a lightly doped *n*-type substrate, a thin blanket layer of SiO₂ (the pad oxide) is formed, and a blanket deposition of a thick protecting Si₃N₄ layer follows.
- Step 2. A resist is deposited on the Si₃N₄ and patter (mask 1). All Si₃N₄ and SiO₂ not covered by the resist (the resist covers the final transistor area) is etched away.
- Step 3. Boron is ion implanted around the perimeter of the resist-protected area to form a *p*⁺-type isolation border (the channel stopper or chanstop). The boron cannot penetrate through the resist.
- Step 4. A thick SiO₂ layer (the field oxide) is grown over the *p*⁺-chanstop to isolate the device. This also drives the *p*⁺-region deeper into the substrate.
- Step 5. The remaining Si₃N₄, pad oxide, and resist are stripped away and a thin, precisely controlled SiO₂ gate oxide layer is thermally grown.

- Step 6.* The doping concentration of the p^+ -type substrate under the gate oxide is adjusted by another boron implantation. Boron passes through the thin gate oxide. This provides the threshold voltage adjustment for the final device.
- Step 7.* A blanket layer of polysilicon and pattern (mask 2), such that the resist covers only the polysilicon that is to become the gate, is deposited. All exposed polysilicon is etched away.
- Step 8.* The n^+ -type source and drain regions are created by As ion implantation. The As can penetrate the thin gate oxide, but not the thick field oxide or the polysilicon gate. The formation of the source and gate does not require a separate resist pattern, thus this technique is called self-aligning.
- Step 9.* SiO_2 is blanket deposited over the substrate. The resist (mask 3) that has openings over the SiO_2 is deposited and patterned. The exposed SiO_2 is etched down to the source, drain, and gate layers, creating contact windows for metallization.
- Step 10.* The system is metallized, first with a tungsten layer, then with Al. The resist is applied and patterned (mask 4), and unwanted metal is etched away.
- Step 11.* If no additional metallization layers are required, the substrate is covered with a passivation layer. If additional levels of metallization are to be added to the structure, a blanket layer of an intermetal dielectric (IMD) is deposited. The resist is deposited, patterned (mask 5), and vias down to the Al in the first metal layer are etched. Steps 10 and 11 are repeated to form the second metal layer.

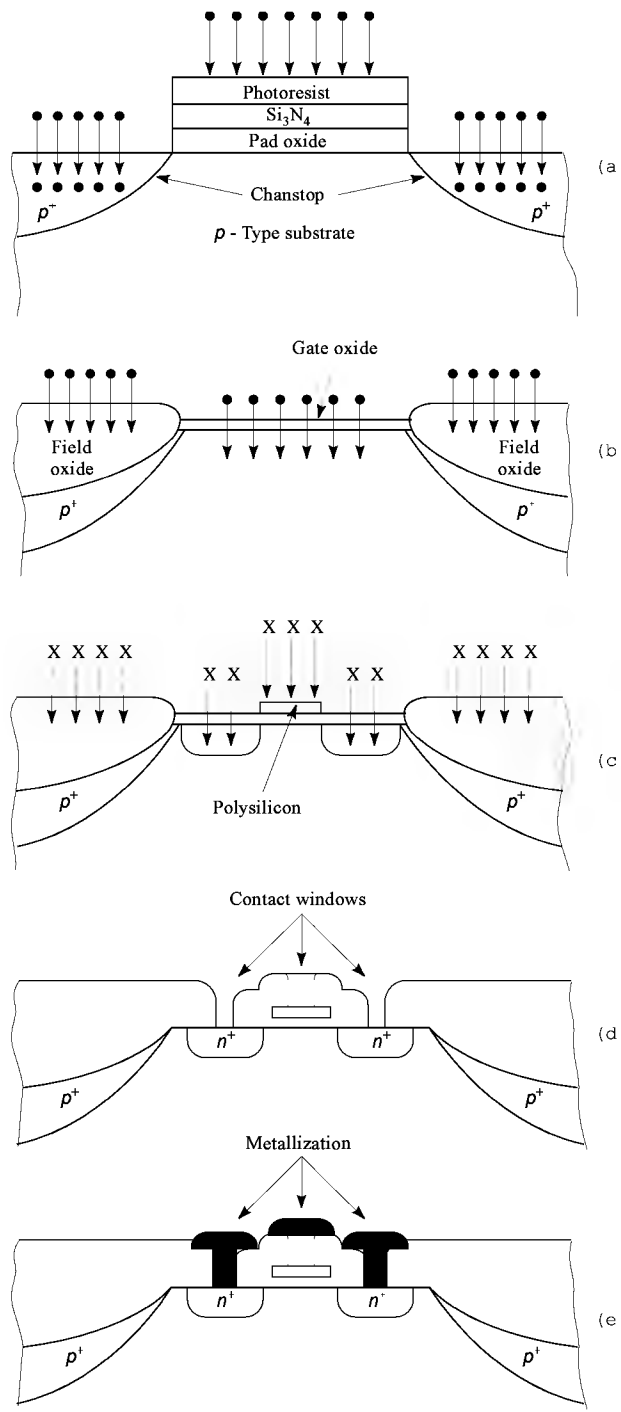


Fig. 9. Fabrication sequence for an oxide-isolated p -well CMOS process, where $\bullet$ is boron and X is arsenic. See text. (a) Formation of blanket pad oxide and Si_3N_4 layer; resist patterning (mask 1); ion implantation of channel stoppers (chanstop) (steps 1–3). (b) Growth of isolation field oxide; removal of resist, Si_3N_4 , and pad oxide; growth of thin ($<200\text{ nm}$) SiO_2 gate oxide layer (steps 4–6). (c) Deposition and patterning of polysilicon gate; formation of n^+ -source and drain (steps 7,8). (d) Deposition of thick SiO_2 blanket layer; etch to form contact windows down to source, drain, and gate (step 9). (e) Metallization of contact windows with W; blanket deposition of Al; patterning of metal (steps 10,11). The deposition of intermetal dielectric or final passivation layer is not shown.

These processes are considerably more complex in actual CMOS fabrication. First, the lower layers of a CMOS structure typically have a twin-tub design which includes both PMOS and NMOS devices adjacent to each other (see Fig. 3b). After step 1, a mask is opened such that a wide area is implanted to form the n -well, followed by a similar procedure to create the p -well. Isolation between active areas is commonly provided by local oxidation of silicon (LOCOS), which creates a thick field oxide. A narrow strip of lightly doped drain (LDD) is formed under the edges of the gate to prevent hot-carrier induced instabilities. Passivation sidewalls are used as etch resists. A complete sequence of fabrication from wafer to packaged unit is shown in Figure 10.

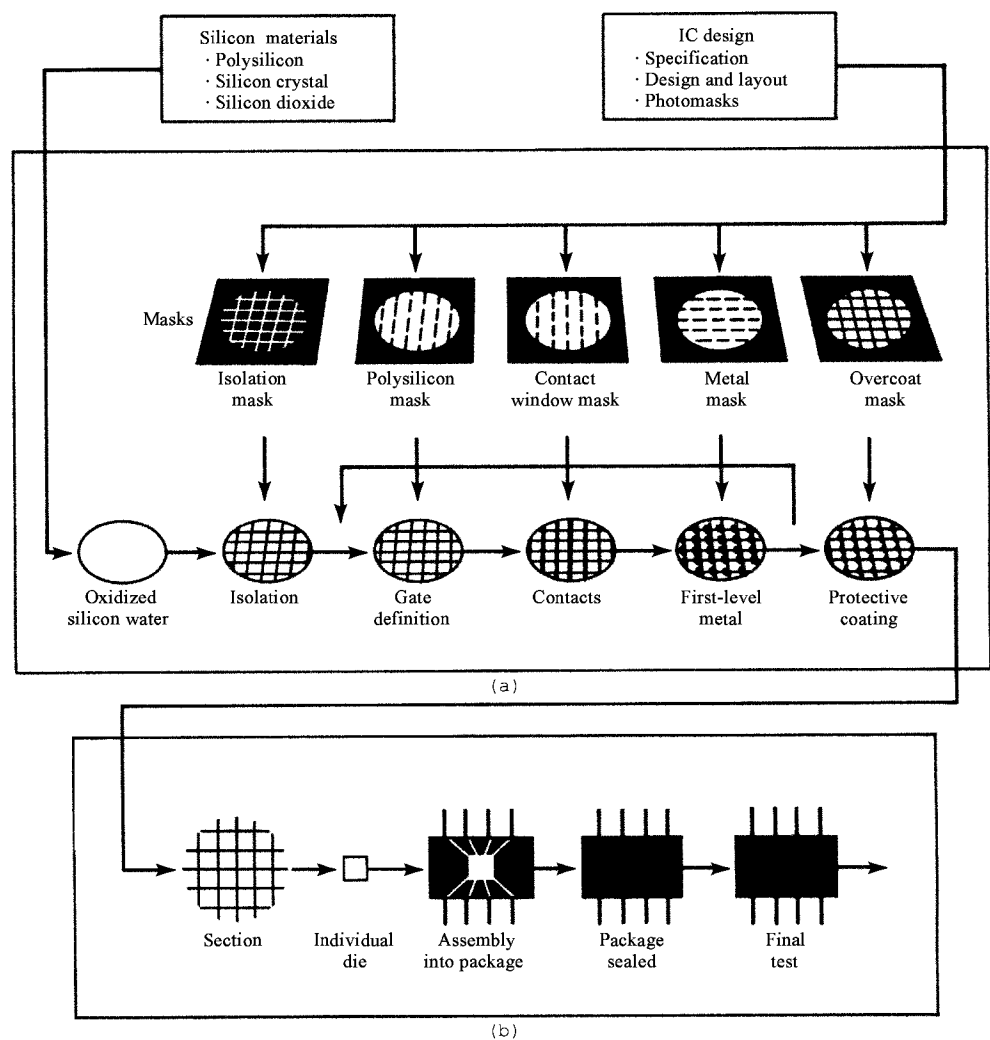


Fig. 10. Complete fabrication sequence for manufacturing a moderately complex silicon device. (a) Front end processing, and (b) assembly and test.

Processing Facilities. Two factors essential to successful IC processing are the quality of the process water and the ambient air. Use of ultrapure filtration (see ULTRAFILTRATION) and clean-room facilities addresses these concerns. In addition, three areas of facilities management are undergoing intense development to support the production of submicrometer VLSI and ULSI designs. The first is the area of particle control, where the thrust is to reduce even further the typical allowance of 0.05 particles cm^2 that are less than 0.35 μm in size (44). There are two approaches to reduce the particle count: prevent particle generation, and prevent particles from depositing on wafers. Both require *in situ* particle characterization technology that is more advanced than what is available as of this writing, where particles are monitored primarily in the vacuum exhaust yielding only secondary information about the cleanliness of a process (32). Particle contamination is most prevalent during rapid changes in the environment of a wafer, such as during vacuum transition (4). Face-down wafer processing has been introduced based on the observation that fewer particulates settle on a wafer surface positioned this way (46). Room air ionizers have also been found to reduce particle counts (47).

The minienvironment approach to contamination control has been increasing in use. A minienvironment is a localized environment created by an enclosure that isolates the product wafer from contamination and people (48). Another approach is using integrated processing, where consecutive processes are linked in a controlled environment (32). Both require *in situ* sensors (qv) to measure internal chamber temperatures, background contamination, gas flow rates, pressure changes, and particularly wafer temperature (4).

A second area of development that has impacted facility design is the trend to single-wafer processing, allowing enhanced control in processing individual wafers. This should carry greater importance as wafer size goes beyond 200-mm diameter to 300–400 mm.

A third facet of facilities development receiving much attention is the concept of total cost of ownership in evaluating process technology and equipment. Cost of ownership goes beyond simple comparison of technological capability and throughput to consider all aspects of equipment and processing costs, including capital costs, final yields, downtime, maintenance (qv) and repair, cost of consumables, and clean-room requirements.

Analytical Techniques

Analytical methodology has had to respond to the rapid scaling of IC designs to submicrometer geometries and to the transition to wafers that are up to 200 mm and larger in diameter. The key requirements that have emerged are not only the need to probe 0.5 μm or smaller features, but also to characterize thin films (qv) that are less than 5 nm thick, detect surface metals present at less than 10^{10} atoms cm^{-3} , and identify trace organic contaminants (see TRACE AND RESIDUE ANALYSIS) (49). These capabilities are essential in process development, process control (qv), and failure mode analysis (FMA), for all stages of fabrication. The purity of gas and liquid starting materials must also be determined (see also NONDESTRUCTIVE EVALUATION; SURFACE AND INTERFACE ANALYSIS).

The physical techniques used in IC analysis all employ some type of primary analytical beam to irradiate a substrate and interact with the substrate's physical or chemical properties, producing a secondary effect that is measured and interpreted. The three most commonly used analytical beams are electron, ion, and photon x-ray beams. Each combination of primary irradiation and secondary effect defines a specific analytical technique. The IC substrate properties that are most frequently analyzed include size, elemental and compositional identification, topology, morphology, lateral and depth resolution of surface features or implantation profiles, and film thickness and conformance. A summary of commonly used analytical techniques for VLSI technology can be found in Table 3.

Table 3. Analytical Techniques Used in VLSI Technology

Energy range, keV	Secondary signal	Acronym	Technique	Application
0.020–0.200	electron	leed	<i>Electron beam</i> low energy diffraction	surface structure

0.300–30	electron	sem	scanning electron microscope	surface morphology
1–30	x-ray	emp	electron microprobe	surface region composition
500–10	electron	aes	Auger spectroscopy	surface layer composition
100–400	electron	tem	transmission electron microscopy	high resolution structure
100–400	electron, x-ray	stem	scanning TEM	imaging, x-ray analysis
100–400	electron	eels	electron energy loss spectroscopy	local small area composition
<i>Ion beam</i>				
0.5–2.0	ion	iss	ion scattering spectrometry	surface composition
1–15	ion	sims	secondary ion mass spectrometry	trace composition vs depth
1–15	atoms	snms	secondary neutral mass spectrometry	trace composition vs depth
>1	x-ray	pixe	particle induced x-ray emission	trace composition
5–20	electron	sim	scanning ion microscope	surface characterization
> 1,000	ion	rbs	Rutherford back-scattering	composition vs depth
<i>Photon beam</i>				
>1	x-ray	xrf	x-ray fluorescence	composition (µm depth)
>1	x-ray	xrd	x-ray diffraction	crystal structure
>1	electron	esca.xps	x-ray photoelectron spectroscopy	surface composition
laser	ion		laser microprobe	composition of irradiated area
laser	light	lem	laser emission microprobe	trace elements (semiquantative)
<i>Neutron beam</i>				
reactor	gamma	naa	neutron activation analysis	bulk (trace) composition

Electron Beam Techniques. One of the most powerful tools in VLSI technology is the scanning electron microscope (sem) (see MICROSCOPY). A sem is typically used in three modes: secondary electron detection, back-scattered electron detection, and x-ray fluorescence (xrf). All three techniques can be used for nondestructive analysis of a VLSI wafer, where the sample does not have to be destroyed for sample preparation or by analysis, if the sem is equipped to accept large wafer-sized samples and the electron beam is used at low (ca 1 keV) energy to preserve the functional integrity of the circuitry. Samples that do not diffuse the charge produced by the electron beam, such as insulators, require special sample preparation.

In the secondary electron mode, a 1–20 keV electron beam is rastered across a surface, causing low (<50 eV) energy electrons to be emitted from the surface to produce a high magnification, high resolution image of a surface. The wide depth of field yields three-dimensional images that are focused even with a wide variation of surface feature heights. The back-scattered mode is used to obtain images with a better contrast between elements of differing atomic number. Sems are also typically equipped for xrf analysis, where a primary x-ray beam generates fluorescent x-rays that are analyzed to qualitatively or semiquantitatively identify the elemental composition of the surface, or to map the distribution of elements in the surface.

A state-of-the-art sem uses a field-emission (fe) electron gun (fesem) to obtain 0.7-nm lateral resolution, equivalent to the resolution obtained by transmission electron microscopy (tem) but with easier and quicker sample preparation (49). Dramatic improvements in image quality have resulted from the development of immersion lenses; 1.0-nm resolution is possible using small (2 mm) samples in off-line inspection (50). High (50–200 keV) energy beams are used to take three-dimensional measurements of high aspect ratio (up to 10:1) contact holes having good depth of field focus. These high energy sources have a sampling depth of up to 20 µm, and can yield images of layers that are covered by other layers, useful for nondestructive detection of voids in underlying films. Lastly, sems are used in nondestructive, on-line, critical dimension (CD) metrology control, where precision as well as speed are required for monitoring dimensional tolerances during processing.

Auger spectroscopy uses a primary electron beam to generate valence-shell electrons that are analyzed for elemental identification and compositional analysis. The two distinguishing features of Auger spectroscopy are that (1) it is very surface sensitive with small spatial resolution. The surface sensitivity stems from the shallow escape depth (0.5–1.0 µm) of the detected electrons, resulting in the characterization of only the uppermost 1–10 monolayers of the surface. (2) In general, Auger spectroscopy can provide quantitative analyses of films as thin as 1.5 nm in areas as small as 15 nm, detecting oxygen and carbon down to 0.1 atomic %. The beam can also be rastered over the surface to provide compositional maps.

Auger spectroscopy is frequently combined with sputter etching to reveal the composition of a surface as a function of depth (depth profile). Used in analysis of ion implantation as well as other applications, the sputter etching removes surface layers to expose underlying layers to the Auger beam. A tenfold improvement in depth resolution has been obtained with the development of the Zalar rotation stage, which reduces crater bottom roughening induced by the ion beam bombardment by rotating the sample during etching (49).

Transmission electron microscopy (tem) is used to analyze the structure of crystals, such as distinguishing between amorphous silicon dioxide and crystalline quartz. The technique is based on the phenomenon that crystalline materials are ordered arrays that scatter waves coherently. A crystalline material diffracts a beam in such a way that discrete spots can be detected on a photographic plate, whereas an amorphous substrate produces diffuse rings. Tem is also used in an imaging mode to produce images of substrate grain structures. Tem requires samples that are very thin (10–50 nm) sections, and is a destructive as well as time-consuming method of analysis.

X-Ray Radiation. After xrf, the most common method of surface analysis utilizing x-rays as the primary beam is x-ray photoelectron spectroscopy (xps), also known as electron spectroscopy for chemical analysis (esca). The x-rays generate photoelectrons emitted with energies that are equal to the difference between the incident photon energy and the binding energy of the electron, and are characteristic for a given element. Xps is very surface-sensitive, similar to Auger analysis, and is particularly useful in analyzing insulators and surface hydrocarbon contaminations. Xps is also used to obtain information about the chemical structure of a film, as the binding energy of an electron is affected by the electronegativity of the surrounding bonded atoms. For example, the carbon 1s signal varies in energy maximum and peak shape depending on the electronegativity of the surrounding bonding atoms. A shortcoming of the technology is the large spot size of the x-ray beam, limiting the lateral resolution of the technique; newer instruments have beam sizes of ca 20 nm.

Ion Beam Techniques. Secondary-ion mass spectrometry (sims) uses ion beams having high enough energies to penetrate the surface and break surface bonds, ejecting neutral and ionic species from the surface in a process called sputtering. The primary beam is typically O⁺. The ejected secondary ions are analyzed and identified according to mass. The sensitivities of ejection, or the ratio of ejected ions to atoms present in the substrate varies greatly according to the particular element, the substrate chemistry, or the substrate structure. The principal advantage of sims analysis is in its very low detection limits, which are applied to the analysis of doping profiles and the detection and identification of surface contaminants. Time-of-flight secondary mass spectrometry (tof-sims) is a new surface-sensitive technique that analyzes both organic and inorganic contaminants in the top monolayer of a surface at ppm detection limits (49). Sims is a destructive analytical process, and requires a large surface area for analysis (5 × 5 mm) (51) (see MASS SPECTROMETRY).

Rutherford back-scattering (RBS) is used to determine the composition and distribution of heavy elements in thin films composed of light elements. ⁴He⁺ ions having energies >1 MeV are accelerated toward a surface, penetrate the surface, undergo elastic collisions with substrate nuclei, and back-scatter out of the surface for detection and analysis. The energy distributions of these ions are characteristic of the elements in the substrate and the depth within the surface where the collisions occur. RBS analysis is nondestructive, and requires a large sample size of several mm. Using hydrogen forward scattering (hfs), it is possible to quantitatively profile hydrogen concentrations as low as 0.1 atomic % to ±10% accuracy on both conducting and insulating samples (49).

Newer techniques that are responding to the need for atomic level imaging and chemical analysis include scanning tunneling microscopes (STMs), atomic force microscopes (AFMs) (52), and focused ion beams (FIBs). These are expected to quickly pass from laboratory-scale use to in-line monitoring applications for 200-mm wafers (32).

BIBLIOGRAPHY

"Integrated Circuits" in *ECT* 3rd ed., Vol. 13, pp. 621–648, by E. F. Labuda and J. T. Clemens, Bell Telephone Laboratories, Inc.

1. C. Hu, *Semiconductor Int.*, 105–114 (June 1994).
2. P. Singer, *Semiconductor Int.*, 56–59 (Apr. 1994).
3. P. Burggraaf, *Semiconductor Int.*, 48–50 (Oct. 1992).
4. M. Moslehi and co-workers, *Solid State Technol.*, 35–45 (Jan. 1994).
5. W. Knausenberger, presentation at the *IPC Technology Marketing Research Council*, New Orleans, La., Dec. 1993.
6. R. Iscoff, *Semiconductor Int.*, 60–66 (Mar. 1992).
7. P. H. Singer, *Semiconductor Int.*, 56–60 (Apr. 1992).
8. P. H. Singer, *Semiconductor Int.*, 44–48 (Jan. 1993).
9. C. W. Pearce, in S. M. Sze, ed., *VLSI Technology*, 2nd ed., McGraw-Hill Book Co., Inc., New York, 1988, p. 10.
10. Ref. 9, p. 59.
11. Ref. 9, p. 63.
12. L. E. Katz, in Ref. 9, p. 125.
13. *Ibid.*, p. 99.
14. *Ibid.*, pp. 108–109.
15. *Ibid.*, p. 113.
16. H. W. Fry and co-workers, *Solid State Technol.*, 31–40 (Mar. 1994).
17. A. C. Adams, in Ref. 9, p. 249.
18. R. J. Schultz, in Ref. 9, p. 225.
19. S. P. Muraka, in Ref. 9, p. 375.
20. *Ibid.*, p. 400.
21. P. Singer, *Semiconductor Int.*, 57–64 (Aug. 1994).
22. P. Singer, *Semiconductor Int.*, 36–39 (Aug. 1992).
23. A. Jain and co-workers, *Semiconductor Int.*, 128–131 (June 1993).
24. J. C. C. Tsai, in Ref. 9, p. 272.
25. S. L. Hurst, *Custom VLSI Microelectronics*, Prentice Hall, Inc., Englewood Cliffs, N.J., 1992, p. 4.
26. Ref. 24, p. 275.
27. Ref. 24, p. 293.
28. M. D. Giles, in Ref. 9, p. 327.
29. *Ibid.*, p. 349.
30. *Ibid.*, p. 345.
31. *Ibid.*, p. 355.
32. Editorial board, *Semiconductor Int.*, 31–40 (Jan. 1993).
33. L. Peters, *Semiconductor Int.*, 82–87 (June 1993).
34. R. K. Watts, in Ref. 9, p. 142.
35. P. Burggraaf, *Semiconductor Int.*, 42–47 (Feb. 1992).
36. P. Burggraaf, *Semiconductor Int.*, 56–59 (Sept. 1994).
37. P. Burggraaf, *Semiconductor Int.*, 52–56 (Mar. 1992).
38. Ref. 34, p. 174.
39. R. J. Schultz, in Ref. 9, p. 193.
40. L. Peters, *Semiconductor Int.*, 66–70 (May 1992).
41. P. Burggraaf, *Semiconductor Int.*, 56–60 (May 1994).
42. P. Singer, *Semiconductor Int.*, 50–54 (Apr. 1993).
43. P. Singer, *Semiconductor Int.*, 78–81 (May 1992).
44. S. J. Fonash, C. R. Viswanathan, and Y. D. Chan, *Solid State Technol.*, 99–107 (July 1994).
45. L. Peters, *Semiconductor Int.*, 66–70 (May 1992).
46. A. Bowling and C. J. Davis, *Solid State Technol.*, 45–50 (Feb. 1994).
47. A. Steinman, *Semiconductor Int.*, 73–75 (Sept. 1994).
48. X. Pucel, *Semiconductor Int.*, 69–74 (Apr. 1994).
49. P. K. Chu and R. S. Hockett, *Semiconductor Int.*, 142–146 (June 1994).
50. L. Peters, *Semiconductor Int.*, 50–53 (Mar. 1994).
51. J. B. Bindell, in Ref. 9, p. 545.
52. B. S. Meyerson, *Solid State Technol.*, 53–57 (Feb. 1994).

Irene Sawchyn, Ph.D.
Consultant

INTEGRATED PEST MANAGEMENT SYSTEMS.

See INSECT CONTROL TECHNOLOGY; REPELLENTS

IODINE AND IODINE COMPOUNDS

Iodine

Iodine [7553-56-2], I, atomic number 53, atomic weight 126.9044, is a nonmetallic element belonging to the halogen family in Group 17 (VIIA) of the Periodic Table. The only stable isotope has a mass number of 127. There are 22 other iodine isotopes having masses between 117 and 139; 14 of these isotopes yield significant radiation.

Iodine was discovered by Curtois in 1811–1812, when he observed violet vapors rising upon heating saltpeter pots (1). Following its discovery and examination, the new element was named *iodé* in French after the Greek word *ioeides* meaning violet-colored. The English term iodine comes from the same root (2).

Iodine is a bluish black, crystalline solid having a metallic luster. It is obtained in shiny flakes or prills that can be easily crushed to powder. Iodine crystallizes in rhomboidal plates belonging to the triclinic system.

Occurrence in Nature. About 99.6^o % of the earth's mass results from 32 of the chemical elements. The remaining 0.4^o % is apportioned among 64 elements, all of which are present as traces. Iodine is one of these 64. Estimates about abundance of the constituent elements of the lithosphere place iodine 46th on a restricted list of 59 elements (37 very rare elements are excluded) and 61st on a list in which 96 elements are included. Iodine is, indeed, one of the scarcest of the nonmetallic elements in the total composition of the earth (3).

Although not abundant in quantity, iodine is distributed in rocks, soils, waters, plants, animal tissues, and foodstuffs (3,4). Excepting the possible occurrence of elemental iodine vapor in the air near certain iodine-rich springs, iodine never occurs free in nature. It is always found combined with other elements.

Wherever it occurs, the quantities of iodine are generally exceedingly small, and very sophisticated chemical methods are required to detect them. Only a few substances characteristically contain iodine in relatively large quantities. These are seaweeds, sponges, and corals (5); the underground waters from certain deep oil-well boring and mineral springs (6); and, most impressive of all, the vast natural deposits of sodium nitrate, caliche, ore found in the northern part of Chile. Even in these, however, the proportion of iodine is small, rarely exceeding one part in five hundred (7–9). Minerals containing iodine as an essential component do exist, but except for those in the Chilean deposits, these minerals are of very rare occurrence (10). Examples are coccinite, Hg₂I₂; cuproiodargyrite, (Ag,Cu)I; iodembolite, Ag(Br,I,Cl); iodobromite, 2AgCl·2AgBr·2AgI; iodagyrite [20704-12-5], AgI; marshite [24401-69-2], Cu₂I₂; miersite, 4AgI·CuI; salesite, CuIO₃(OH); schwarzembergite, Pb(ICI)₂·2PbO; tocornalite, (Ag,Hg)I; dietzeite, 7Ca(IO₃)₂·8CaCrO₄; and lautarite [7789-80-2], Ca(IO₃)₂. Lautarite and dietzeite are the two crystalline forms in which iodine naturally occurs in caliche, the natural saltpeter Chilean mineral (11).

Physical Properties. The absorption of x-rays by iodine has been studied and the iodine crystal structure determined (12,13). Iodine crystallizes in the orthorhombic system and has a unit cell of eight atoms arranged as a symmetrical bipyramid. The cell constants at 18°C (14) are given in Table 1, along with other physical properties. From the interatomic distances of many iodine compounds, the calculated effective radius of the covalently bound iodine atom is 184 pm (15).

Table 1. Physical Properties of Iodine

Properties	Solid	Liquid	Gaseous
color	bluish black	bluish black	violet
melting point	113.6		
boiling point, °C		185	
critical temperature, °C		553	
critical pressure, kPa ^c		11753.7	
density, g mL			
20°C	4.93		
60°C	4.89		
120°C		3.960	
180°C		3.736	
vapor density at 101.3 kPa, ^d g L			6.75
cubic coefficient of expansion 0–113.6°C, °C ⁻¹	2.81 × 10 ⁻⁴		
crystal structure, 4 mol I ₂ per unit cell	orthorhombic		
unit cell dimensions at 18°C, pm			
<i>a</i>	477.61		
<i>b</i>	725.01		
<i>c</i>	977.11		
entropy at 25°C, J (molK) ^e	116.81		62.25
specific heat, J (gK) ^e			
25–113.6°C	0.1582 + 1.9628 × 10 ⁻⁴ T ^f		
113.6–184°C		0.3165	
25–1200°C			0.1465
heat of fusion at 113.6°C, J g ^g	62.17		
heat of sublimation at 113.6°C, J g ^g	238.40		

heat of vaporization at bp, J g ^e		164.45
viscosity, mPa(cP)		
at 116°C		2.268
at 185°C		1.414
vapor pressure, kPa ^d		
at 25°C	0.04133 ^a	
at 113.6°C	12.0655	^b
thermal conductivity at 24.4°C, W (m·K)	0.4581	
electrical resistivity, Ωcm		
at 25°C	5.85 × 10 ⁶	
at 110°C	8.33 × 10 ⁶	
at 140°C		1.1 × 10 ⁵
dielectric constant		
at 23°C	10.3	
at 118°C		11.08
refractive index, n _D	3.34	

^a For the solid, between 0 and 113.6°C, log *p*_{kPa} = (3410.71 / *T*) – 0.3523 log *T* – 1.301 × 10^{–3} *T* + 14.3140 where *T* is in Kelvin.

^b For the liquid, between 113.6 and 186°C, log *p*_{kPa} = (2300.24 / *T*) + 10.025, where *T* is in Kelvin.

^c To convert kPa to atm, divide by 101.

^d To convert kPa to mm Hg, multiply by 7.50.

^e To convert J to cal, divide by 4.184.

^f *T* is in Kelvin.

The density, *d*, of liquid iodine (16) is given by the expression

$$d = 3.94916 + 0.003267 \times (120 - t)$$

(1)

where *t* is temperature in °C. The specific volume at 120°C is 0.2532 mL. Measurements of the density of solid iodine have been extended to low temperatures, eg, *d* at –195°C = 5.15 g / mL; the atomic volume at –273°C obtained by extrapolation is 24.5 mL (17,18). The experimental values for the viscosity of iodine vapor at 8 kPa (60 mm Hg), are determined over the temperature range 380–800°C, in reasonable agreement with those calculated by the Sutherland equation (19,20). The viscosity of liquid iodine at a temperature just above the melting point is 1.98 mPa(cP) as compared to the 2.35 mPa (cP) calculated by Andrade and Lindemann's equation. The coefficient of cubic expansion of solid iodine has been found to be 2.64 × 10^{–4} mL / °C, and that for liquid iodine 8.54 × 10^{–4} mL / °C (21,22). The critical temperature, *T*_c of iodine has been determined as 553°C (23); the most probable value for the melting point of iodine is 113.6°C (24,25), although it has been claimed that a prolonged drying may raise it to 115.6°C (26). The change in volume on fusion is 2.04° o.

The vapor pressure, *p*, of solid iodine has been redetermined using the gas current method and by a static method using a flexible metallic diaphragm (27,28). The data from the gas current method are well represented by equation 2 (27):

$$\log p_{\text{kPa}} = (3512.8 / t) - 2.013 \log t + 15.3796$$

(2)

Temperature, °C	Vapor pressure, Pa (psi)
0°C	3.97 (0.0006)
10°C	10.64 (0.0015)
20°C	26.50 (0.0038)
30°C	61.90 (0.0090)
40°C	136.46 (0.0198)
50°C	285.56 (0.0414)
60°C	570.35 (0.0827)
70°C	1090.44 (0.1581)
80°C	2006.10 (0.2909)
90°C	3561.03 (0.5163)
100°C	6118.15 (0.8871)

The diffusion coefficient of iodine vapor between 14 and 30°C varies between 0.0767 to 0.0851 (29). The latent heat of fusion of molecular iodine has been found to be 15.66 ± 0.08 kJ / mol (3.74 ± 0.02 kcal / mol), and the latent heat of vaporization, calculated from vapor pressure data, to be 62.30 kJ / mol (14.89 kcal / mol) (28,30).

The heat capacity of iodine at constant pressure has been redetermined for the temperature ranges –263 to –223°C (30) and –71 to 160°C (31). For solid iodine, in the interval 25–113.6°C, it can be expressed by equation 3:

$$C_p = 3.122 + 0.781 \times 10^{-4} \times (T - 298)^2$$

(3)

where *C*_p is the heat capacity in J / (mol·K) and *T* is the temperature in K (31). The heat capacities of iodine above its melting point are available (32). For the interval 113.6 to 160°C, a constant molal heat capacity for liquid iodine of 81.64 J / (mol·K) 19.51 kcal / (mol·K) is reported.

The thermal conductivity of solid iodine between 24.4 and 42.9°C has been found to remain practically constant at 0.004581 J / (cm·s·K) (33). Using the heat capacity data, the standard entropy of solid iodine at 25°C has been evaluated as 116.81 J / (mol·K), and that of the gaseous iodine at 25°C as 62.25 J / (mol·K), which compares satisfactorily with the 61.81 value calculated by statistical mechanics (34,35).

Iodine is only slightly soluble in water and no hydrates form upon dissolution. The solubility increases with temperature, as shown in Table 2 (36). Iodine is soluble in aqueous iodide solutions owing to the formation of polyiodide ions. For example, an equilibrium solution of solid iodine and KI₇·H₂O at 25°C is highly concentrated and contains 67.8° o iodine, 25.6° o potassium iodide, and 6.6° o water. However, if large cations such as cesium, substituted ammonium, and iodonium are present, the increased solubility may be limited, owing to precipitation of sparingly soluble polyiodides. Iodine is also more

soluble in solutions of chlorides, bromides, and other salts than in pure water, but these salts have much less effect than do the iodides (36).

Table 2. Solubility in Water^a

Temperature, °C	Solubility, g kg	Temperature, °C	Solubility, g kg
0°C	0.162	60°C	1.06
20°C	0.293	70°C	1.51
25°C	0.340	80°C	2.17
30°C	0.399	90°C	3.12
40°C	0.549	100°C	4.48
50°C	0.769	110°C	6.65

^a Ref. 36.

Iodine dissolves in many organic solvents (Table 3) (36), and the color of the resulting solutions varies with the nature of the solvent. This color variation results from the particular charge transfer possible upon complex formation. Complexation occurs with solvents that are electron donors. Aliphatic hydrocarbons and carbon tetrachloride give violet solutions, aromatic hydrocarbons give pink or brown-pink solutions, and alcohols, ethers, and high vapor amines give brown colors. Iodine readily sublimes, as is evident by the pressure below the melting point. It also exists as a black mobile liquid at atmospheric pressure over a range of more than 60°C. Iodine, a good solvent for alkali metal iodides and ammonium bases, dissolves sulfur, selenium, the covalent iodides of such metals as aluminum, tin, and titanium, and many organic substances.

Table 3. Solubility in Organic Solvents^a

Solvent	Solubility, g kg
benzene	164.0
carbon disulfide	197.0
carbon tetrachloride	19.2
chloroform	49.7
cyclohexane	27.9
ethyl acetate	157.0
ethyl alcohol	271.1
ethyl ether	337.3
ethylene bromide	115.1
ethylene chloride	57.6
glycerol	9.7
<i>n</i> -heptane	17.3
<i>n</i> -hexane	13.2
isobutyl alcohol	97
isooctane	13.2
tetrachloroethylene	69 ^b
toluene	182.5
trichloroethylene	79 ^b
<i>p</i> -xylene	198.3
water	0.34

^a At 25°C. Ref. 36.

^b Solubility is given in g 1000 mL. of solution.

Solutions of alkali metal and ammonium iodides in liquid iodine are good conductors of electricity, comparable to fused salts and aqueous solutions of strong acids. The liquid is therefore a polar solvent of considerable ionizing power, whereas its own electrical conductivity suggests that it is appreciably ionized, probably into I⁺ and I₃⁻ (triiodide). Iodine resembles water in this respect. The metal iodides and polyiodides are bases, whereas the iodine halides are acids.

Iodine vapor is characterized by the familiar violet color and by its unusually high specific gravity, approximately nine times that of air. The vapor is made up of diatomic molecules at low temperatures; at moderately elevated temperatures, dissociation becomes appreciable. The concentration of monoatomic molecules, for example, is 1.4^o % at 600°C and 101.3 kPa (1 atm) total pressure. Iodine is fluorescent at low pressures and rotates the plane of polarized light when placed in a magnetic field. It is also thermoluminescent, emitting visible light when heated at 500°C or higher.

Chemical Properties. The electron configuration of the iodine atom is [Kr]4*d*¹⁰5*s*²5*p*⁵ and its ground state is 2*p*_{1/2}^o. Principal oxidation states are -1, +1, +3, +5, and +7, but the oxide IO₂ where iodine has an oxidation state of +4 is also known. Iodine forms thermodynamically stable compounds in all these oxidation states, except the +4. Iodine is the heaviest of all the common halogens and the least electronegative. It is usually less violent in its reactions than the other members of the halogen family. Iodine presents mild oxidizing properties in acidic solutions.

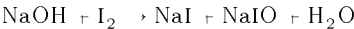
Iodine forms compounds with all the elements except sulfur, selenium, and the noble gases. It reacts only indirectly with carbon, nitrogen, oxygen, and some noble metals such as platinum.

Iodine, as a vapor or as a solid, reacts at room temperatures with the vapors of the alkali metals in a highly luminous reaction (37,38). Iodine as a vapor or in the solid state reacts at room temperature with copper and silver to form the respective iodides. Iodine as a vapor or in the solid state reacts with magnesium, calcium, aluminum, zinc, tin, nickel, and iron, forming the corresponding iodides. It does not react with lead, bismuth, or gold (39). Platinum is not attacked at ordinary temperatures, but reacts readily at 1400°C, forming PtI and PtI₂. Tungsten is attacked by iodine at ordinary temperatures, forming the iodide, but is not attacked at high temperatures, where even atomic iodine has no effect. Iodine reacts with tantalum and niobium at 1300–1500°C (40), where the corresponding pentaiodides are formed. Iodine and vanadium react to form VI₅, which on heating decomposes to V₂I₃, which does not decompose on further heating.

Moist iodine vapor rapidly corrodes metals, including most stainless steels. The initial process is the formation of corrosion centers where small amounts of metal iodide are formed which deliquesce, and the corrosion then takes place electrochemically (41,42). Only titanium and molybdenum steels are unattacked by iodine (42,43). The corrosion of molten iodine has also been studied.

Under high pressures and temperatures, iodine reacts with oxygen to form iodine pentoxide [12029-98-0] (44). The reaction of iodine with carbon monoxide under acidic conditions is catalyzed by palladium salts (45). Phosphorous vapor and iodine react to form phosphorus triiodide [13455-01-1], PI₃ (46).

Iodine dissolves without reaction in concentrated sulfuric acid and with concentrated nitric acid it reacts to form iodine pentoxide (47). Iodine reacts with alkali metal hydroxide solutions to form the corresponding hypoiodite and the rate of the reaction increases with the alkali concentration and temperature. At 50°C, the reaction is almost instantaneous:



Subsequently, the hypoiodite is oxidized to iodate, and this reaction is not influenced by the alkali concentration, temperature, or iodate concentration.

Iodine reacts with hydrocarbons to form iodine compounds, but compared to the other halogens, the equilibria are unfavorable because the displacement step with the iodine atom is endothermic, requiring 4066.3 J (971.9 cal) for methane and 799.9 J (191.2 cal) for toluene. Hydrogen iodide can be used to reduce an alkali iodide to hydrocarbon plus molecular iodine.

Complete iodination of organic compounds can be achieved by preventing the formation of free hydrogen iodide through the addition of an oxidizing agent, neutralization of the HI with a base, or combination using mercuric salts.

Iodine adds to carbon–carbon double bonds to form polyiodine derivatives. These addition reactions are reversible, however, and do not go to completion.

Reactions in Aqueous Media. The chemistry of aqueous iodine has been extensively studied because of the role of iodine as a disinfectant (see DISINFECTANTS AND ANTISEPTICS). The system is very complex, owing to the number of oxidation states available to iodine under ambient conditions (48).

Aqueous solutions of iodine are hydrolyzed according to



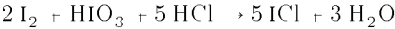
The equilibrium constant of this reaction is 5.4×10^{-13} at 25°C, ie, iodine hydrolyzes to a much smaller extent than do the other halogens (49). The species concentrations are highly pH dependent: at pH = 5, about 99% is present as elemental I₂; at pH = 7, the I₂ and HIO species are present in almost equal concentrations; and at pH = 8, only 12% is present as I₂ and 88% as HIO. The dissociation constant for HIO is ca 2.3×10^{-11} and the pH has little effect on the IO⁻ ion formation. At higher pH values, the HIO converts to iodate ion. This latter species has been shown to possess no disinfection activity. An aqueous solution containing iodate, iodide, and a free iodine or triiodide ion has a pH of about 7. A thorough discussion of the kinetics of iodine hydrolysis is available (49).

Iodine is a mild oxidizing agent in acidic solutions, having an equilibrium potential of the iodine–iodide ion couple of 0.5345 V at 25°C (50,51). Iodine readily oxidizes sulfur dioxide to sulfate, thiosulfate to tetrathionate, and stannous and titanous salts to stannic and titanic salts. On the other hand, ferric and cupric salts, and salts of vanadium, chromium, and manganese in the respective highest oxidation states, oxidize iodide ion in acid solution, liberating free iodine. Oxidizing agents such as chlorine, bromine, nitrous acid, and hot nitric acid liberate iodine from iodide solutions (52).

Iodine can be oxidized to iodate in acid solutions by concentrated nitric acid and, in more dilute solutions, by permanganate, bromates, chlorates, and even chlorine and bromine.



However, in strong hydrochloric acid, these reagents, as well as iodic acid, oxidize iodine to iodine monochloride or to the ICl₂ ion.



In alkaline solutions, iodine can be oxidized to iodate by sodium hypochlorite or hypobromite, whereas chlorine passed into a solution of iodine and alkali oxidizes all the way to periodate.



Hydroiodic acid reacts quantitatively with iodic acid to give iodine:



The kinetics of this reaction, the conditions under which it goes to completion, and subsequent titration of the iodine with thiosulfate solutions have all been investigated (53–55).

Small concentrations of HI reduce concentrated sulfuric acid to sulfurous acid; high concentrations of HI reduce it to hydrogen sulfide (56).

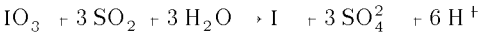
Manufacture and Processing. The industry related to iodine production began a few years after the discovery of the element by Courtois in 1811. The production processes are based on the raw materials containing iodine: seaweeds, mineral deposits, and oil-well or natural gas brines.

Seaweeds. The earliest successful manufacture of iodine started in 1817 using certain varieties of seaweeds. The seaweed was dried, burned, and the ash lixiviated to obtain iodine and potassium and sodium salts. The first process used was known as the kelp, or native, process. The name kelp, initially applied to the ash of the seaweed, has been extended to include the seaweed itself. About 20 t of fresh seaweed was used to produce 5 t of air-dried product containing a mean of 0.38 wt % iodine in the form of iodides of alkali metals. The ash obtained after burning the dried seaweed contains about 1.5 wt % iodine. Chemical separation of the iodine was performed by lixiviation of the burned kelp, followed by solid–liquid separation and water evaporation. After separating sodium and potassium chloride, and sodium carbonate, the mother liquor containing iodine as iodide was treated with sulfuric acid and manganese dioxide to oxidize the iodide to free iodine, which was sublimed and condensed in earthenware pipes (57).

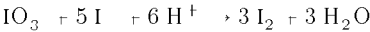
Mineral Deposits. The only iodine obtained from minerals has been a by-product of the processing of nitrate ores in Chile. Caliche occurs in the Atacama desert of Northern Chile and west of the Andes mountains. The Atacama desert is known as the driest of the world's deserts, where measurable (>1 mm) rainfalls may be as infrequent as once every 5–29 years (58). The caliche deposits occur in an area averaging 700 km (north–south) by 30 km (east–west). The iodine may total over 5×10^6 t (59).

The first iodine recovery from caliche occurred in 1852; the first iodine was exported to Europe in 1868, becoming the most important by-product of the nitrate production in terms of value. There are two ways for producing iodine from caliche iodates: first, from solutions containing more equivalent iodine than its solubility as elemental iodine in the same solution of about 0.4 g/L at 25°C; and second, from more diluted equivalent iodine solutions.

First Alternative. Figure 1 illustrates the first of the two alternative production processes. Here the mother liquor from the sodium nitrate crystallization plant, normally containing about 1.5 g/L iodine as iodate, is decanted for clarification and concentration homogenization. From there the solution is split into two fractions. The larger fraction is fed into an absorption tower where it is contacted with SO₂ obtained by sulfur combustion. In the absorption tower iodate is reduced to iodide according to the following reaction:



After leaving the absorption tower, the resulting iodide solution, together with the iodide solution coming from the kerosene extraction plant as described below, is contacted with the smaller iodate fraction in the stoichiometric proportion shown, producing iodine:



Up to 0.4 g/L of the iodine stays in solution and the rest precipitates as crystallized iodine, which is removed by flotation (qv). This operation does not require a flotation agent, owing to the hydrophobic character of the crystallized element. From the flotation cell a heavy pulp, which is water-washed and submitted to a second flotation step, is obtained. The washed pulp is introduced into a heat exchanger where it is heated under pressure up to 120°C to melt the iodine that flows into a first reactor for decantation. From there the melt flows into a second reactor for sulfuric acid drying. The refined iodine is either flaked or prilled, and packed in 50- and 25-kg plastic-lined fiber drums.

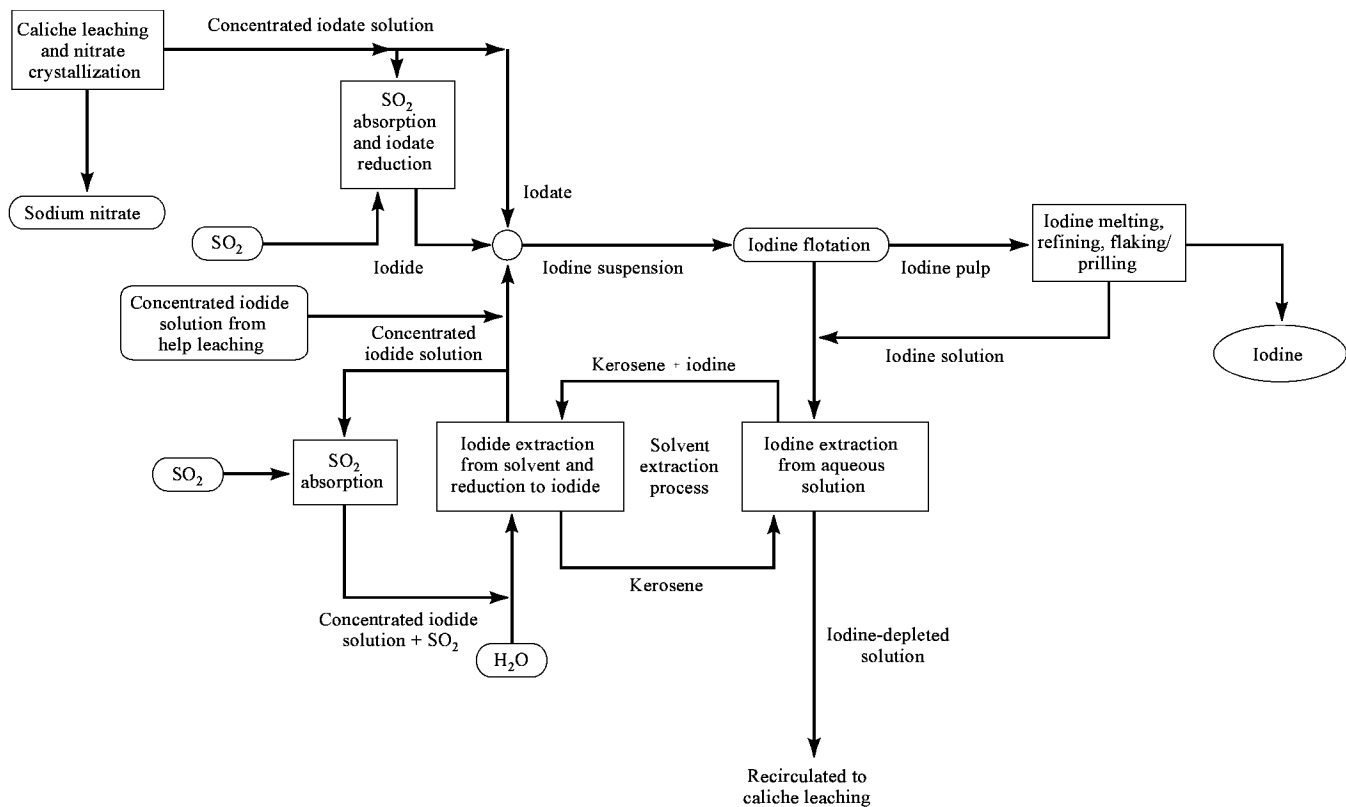


Fig. 1. SQM nitrate iodine production process for obtaining iodine from concentrated iodate solutions.

The solution leaving the flotation cell, containing about 0.4 g L iodine, is sent to a kerosene solvent extraction process to recover the dissolved product. After neutralization with soda ash to the initial incoming alkalinity, the solution is returned to the nitrate lixiviation process. The iodine-charged kerosene is contacted with an acidic concentrated iodide solution containing SO₂, which reduces the iodine to iodide.

Second Alternative. The second alternative production process is shown in Figure 2. The treatment of a diluted iodate solution does not require a flotation step, because all the iodine stays in solution. Therefore only the kerosene extraction unit is used, and the final product of this plant is a concentrated iodide solution, which is used to react with the iodate mother liquor stream of the plants using concentrated iodate solutions.

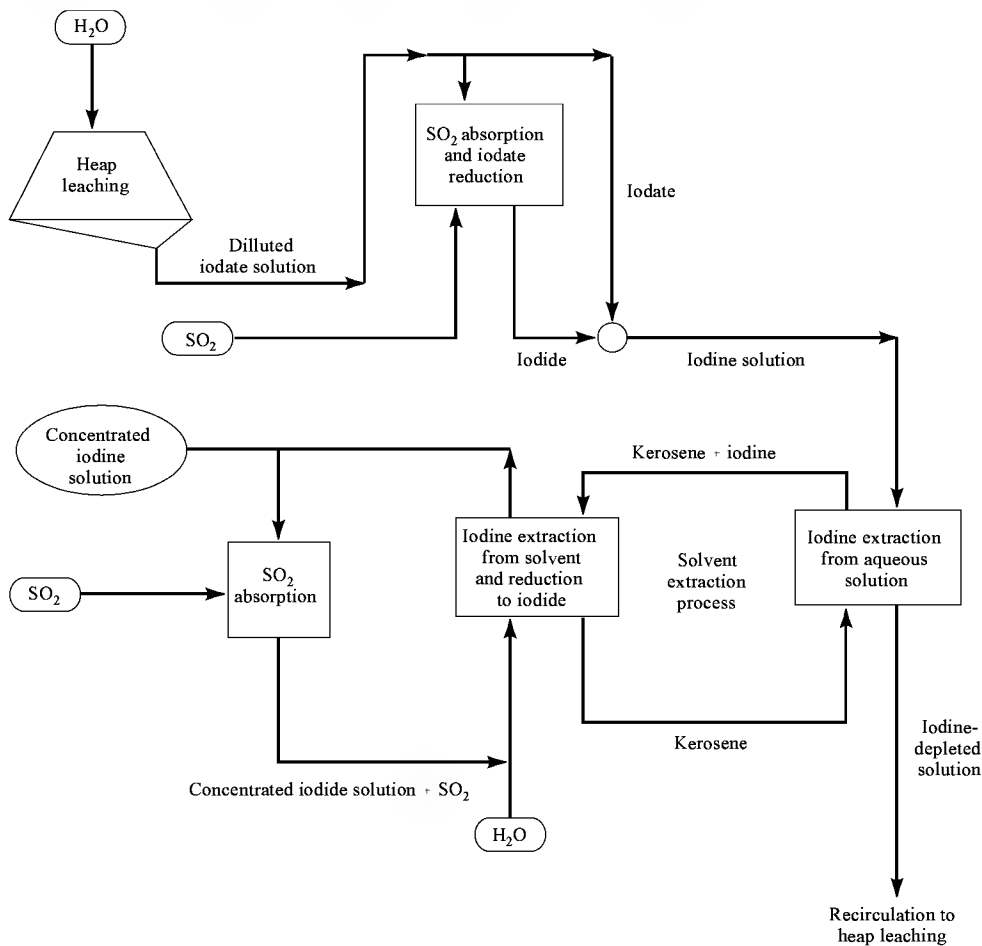


Fig. 2. SQM's heap leaching of caliche tailings for iodine recovery from diluted iodate solutions.

Diluted iodate solution is obtained by hydrothermal vat leaching of caliche ore during nitrate recovery. Concentrated iodide solutions are obtained by heap leaching of old waste dumps (tailings) and low grade nitrate caliche, such as blasted overburden, left over by former nitrate producers.

Until 1990, some flaked iodine and iodine-containing wastes were sublimed to obtain an especially high grade product. This process has been dropped because of quality improvements in the standard operations. Sublimed iodine is produced only in small quantities by specialized companies that

offer this product for special minor applications requiring the highest purity.

Brines. About 65% of the iodine consumed in the world comes from brines processed in Japan, the United States, and the former Soviet Union (see CHEMICALS FROM BRINE). The predominant production process for iodine from brines is the blow-out process, which was first used in Japan. Iodine is present in brines as iodide, and its concentration varies from about 10 to 150 ppm. As shown in Figure 3, the recovery process can be divided into brine clean-up, iodide oxidation to iodine followed by air blowing out and recovery, and iodine finishing.

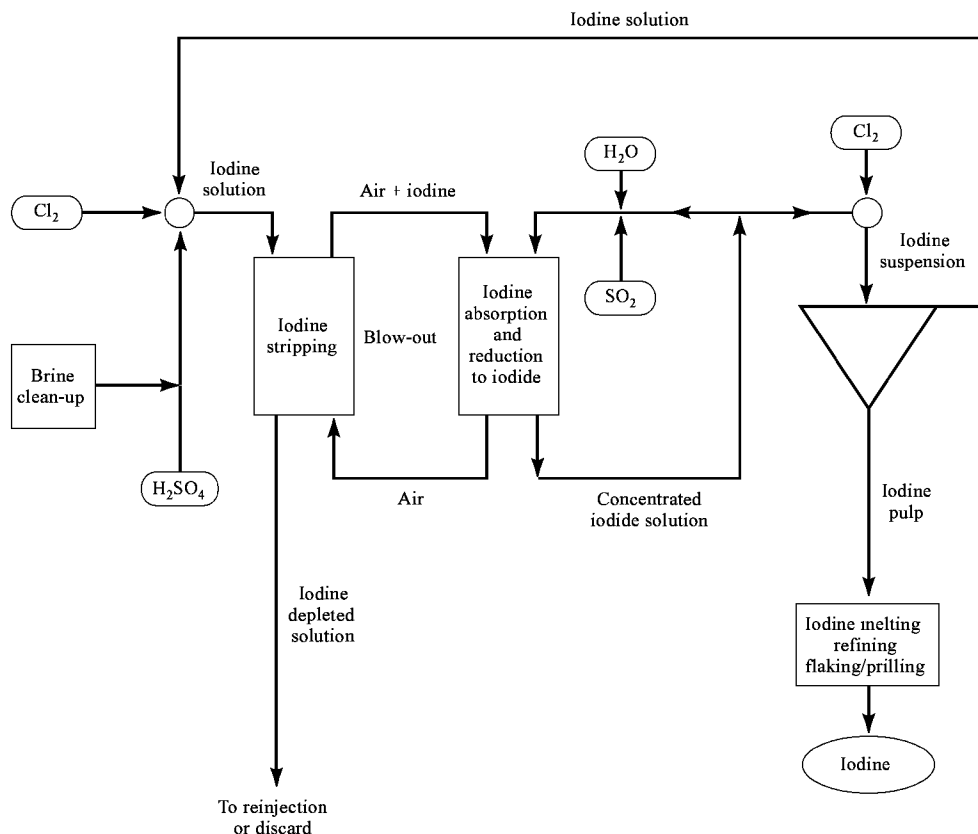
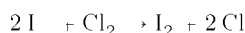


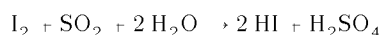
Fig. 3. Iodine production process from brines.

The brine clean-up consists of skimming and settling steps to free the solution from oil, clays, and other impurities. Sulfuric acid is then added until a pH of <2.5 is reached ensuring iodine liberation by oxidation, precipitation of the soluble barium contained in the brine, and recovery of the remaining iodine.

The clarified and acidified brine is treated with gaseous chlorine which is injected into the solution in a small excess over the theoretical stoichiometric relation by weight of 0.28:1 chlorine to iodide. The oxidation occurs according to



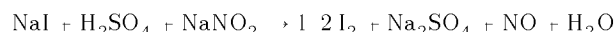
The I_2 formed stays in solution, exerting a certain vapor pressure, and is extracted from the brine in a countercurrent air blow-out process. The extracted brine leaves the extraction tower and is discarded or reinjected into the wells to avoid sinking of the soil. The iodine-loaded air is then submitted to a cocurrent desorption process by means of an acidic iodide solution to which SO_2 is added. By this solution the iodine is reduced to iodide by the following reaction:



Part of the continuously recirculated solution is bled off and sent to the iodine finishing process. Iodine finishing consists of contacting this bleed of concentrated acidic iodide solution with gaseous chlorine, through which iodine is formed by oxidation and precipitated. After iodine precipitation, the resulting acidic mother liquor, saturated with free iodine, is pumped back to acidify the clarified brine and to recover the remaining iodine.

The crystallized iodine is decanted and transferred into a fusion kettle. The melted product is contacted with strong sulfuric acid to remove organic impurities and humidity. Finally the iodine is flaked or prilled and packed.

For brines having very low iodide concentrations, ie, in some facilities in Japan and in the former USSR, the activated carbon method of recovery is used. This method consists of a process involving the treating of the acidified brine with sodium nitrite in large tanks, where the following reaction takes place:



The free iodine, which remains in solution, is recovered by adsorption on activated carbon. After this step is completed, the residual brine is neutralized with lime or ammonia.

The iodine is extracted from the activated carbon using hot caustic soda. The product obtained is an iodate-iodide solution, according to the following reaction:



This solution is treated with H_2SO_4 and $\text{K}_2\text{Cr}_2\text{O}_7$, precipitating the iodine crystals and filtering them out, and pressing the cake to remove impregnation. The rest of the process consists of subliming the cake directly or in submitting it to melting and flaking.

The newest process uses ion-exchange (qv) resins on brines already oxidized to liberate iodine. The liberated iodine in the form of polyiodide is adsorbed on Amberlite IRA-400, an anion-exchange resin. When the ion-exchange resin is saturated, it is discharged from the bottom of the column and then transferred to the elution column. Iodine is eluted (or desorbed) using caustic solution followed by sodium chloride. The regenerated resin is returned to the adsorption column. The elutriant, rich in iodide and iodate ions, is acidified and oxidized to precipitate iodine. The crude iodine is then separated in a centrifuge and purified with hot sulfuric acid or refined by sublimation. A patent for this process was granted in the United States in 1939 (57); the first plants to use this process were located in Japan and went on-stream between 1963 and 1966. In 1991, two plants in Japan produced iodine by the ion-exchange process: Nihon Tennen Gas Kogyo Co. Ltd. (NITTEN) and Kanto Natural Gas Co. (60).

Materials of Construction. High silicon iron, Stellite 6, Hastelloy C, and stainless steels types 304, 309, 316, and 317, have low corrosion

rates when immersed in an aqueous solution containing 5° I₂ and 7.5° KI at 25°C. Other metals and alloys are less satisfactory. A 5° solution of iodine in 95° alcohol is more severe in its attack on the stainless steels, but the rate is reduced by the addition of potassium iodide. Iron and steel are badly corroded under most conditions, but medium carbon steel (0.21° carbon) has an extremely low corrosion rate in a solution of 26.8 g L. of iodine in benzene when air and moisture are excluded (61).

Hastelloy B gives promising results when exposed to liquid and gaseous iodine up to 400°C; molybdenum seems capable of withstanding even higher temperatures. High silicon iron, lead-coated steel, and hard lead have been recommended as materials of construction for the sublimation of iodine (62). Niobium, tantalum, molybdenum, and their alloys are resistant to corrosion in a range from 20 to 270°C in moist iodine, liquid iodine, and gaseous iodine (63).

Among nonmetallic materials, glass, chemical stoneware, enameled steel, acid-proof brick, carbon, graphite, and wood are resistant to iodine and its solutions under suitable conditions, but carbon and graphite may be subject to attack. Polytetrafluoroethylene withstands liquid iodine and its vapor up to 200°C although it discolors. Cloth fabrics made of Saran, a vinylidene chloride polymer, have lasted for several years when used in the filtration of iodine recovered from oil-well brines (64).

Economic Aspects. Iodine traditionally has had boom and bust cycles, but the oversupply scenario appearing by the end of 1990 was more acute than previous lows. Tight supply sent prices up to a peak of \$22 kg in 1988. The price increase encouraged fast expansion of the capabilities of iodine-producing companies. The Chilean producers expanded caliche ore extracting, and Japanese and German-owned companies made investments in U.S. brine-based iodine production. Consequently, iodine prices fell from \$19 kg at the beginning of 1990 to \$12 kg by the end of the same year, reaching \$9.50 kg in 1992. This situation may bring pressure on higher cost brine-based producers, possibly forcing curtailed production (65).

Iodine plant locations in the United States and Japan are dictated primarily by the availability of natural brines or bitterns containing adequate amounts of iodine. In 1992, the United States had three iodine-producing companies: Woodward Iodine Corp., North American Brine Resources, and IoChem. In Japan there are five iodine-producing companies, with over 30 plants: Ise, Godo, Nippo, Nitten, and Kanto. All these companies deliver iodine as flaked material except Ise, which also produces prilled iodine.

Plants in the United States are basically iodine producers and must extract the solutions from deep (between 2000- and 3000-m) wells. The depleted solutions are reinjected for environmental reasons and maintain the pressure of the exploitation area. In Japan, on the other hand, iodine is mainly a by-product of natural gas production, and the wells are less deep (about 1500 m). Depleted solutions are often discarded into the ocean. Costs associated with deep wells are relatively high, reaching \$1.7 to 2.0 × 10⁶ in the United States and up to ca \$0.7 × 10⁶ in Japan.

Plant investment and maintenance costs are relatively high for a new iodine plant in the United States or in Japan because of the deep wells required for brine production and disposal as well as the corrosive nature of the plant streams. The principal materials cost is for chlorine and for sulfur dioxide, although in the United States the additives used for the brines, such as scale inhibitors and bactericides, also have a considerable influence on costs.

For the Chilean iodine, which is mainly associated with nitrate production, plant location is adjacent to the nitrate plants, although some smaller installations are independent and are installed where high level iodate-containing old waste dumps or nitrate tailings are available. In 1992 in Chile there were four iodine-producing companies in operation. The largest, SQM, is the largest world iodine producer. Its production during 1992 was ca 4900 metric tons, accounting for 80° of the Chilean iodine output. Minor Chilean producers are ACF Minera Ltd., Minera Sierras de Tarapaca, and Minera Mapocho. Besides the Japanese company Ise, which is offering prilled iodine to the market, SQM is the only iodine producer that in 1992 developed its own iodine-prilling technology.

The principal material cost for the Chilean iodine producers is sulfur used for the manufacture of sulfur dioxide, the reducing agent for the iodates contained in the leach solutions. Also, the use of sodium carbonate for the neutralization of the depleted solutions is an important cost factor.

Not considering the former USSR, world production of iodine was ca 13,500 metric tons in 1992. Japan provided about 45° of the world total, compared to 44° from Chile and 11° from the United States. An annual output of 2300 t from 1976 to 1979 was estimated by the U.S. Bureau of Mines (66) but was revised to 2000 tons in 1981. No official data are available for the former USSR where iodine production is reported to be produced from iodine–bromine brines. Two areas have been mentioned: the Neftechalinki field in the Slavianski-Triotskoe area near the Black Sea, and a plant in the Baku area in Azerbaidzhan on the Caspian Sea where ca 1400 metric tons was estimated for 1990 production.

Specifications and Standards, Shipping. Commercial iodine has a minimum purity of 99.8°. The Committee of Analytical reagents of the American Chemical Society (67) and the *U.S. Pharmacopoeia XXII* (68) specify an iodine content not less than 99.8°, a maximum nonvolatile residue of 0.01°, and chlorine–bromine (expressed as chlorine) of 0.005° (ACS) and 0.028° (USP), respectively. In the past these requirements were attained basically only by sublimation, but with processing changes these specifications can be met by direct production of iodine. Previously the impurities of the Chilean product were chiefly water, sulfuric acid, and insoluble materials. Improvements in the production process, and especially in the refining step, allow the direct obtainment of ACS-type iodine. Also, because of its origin and production process, the Chilean iodine has a chlorine–bromine impurity level of no more than 0.002°.

Iodine is packed in double polyethylene-lined fiber drums containing 10, 25, and 50 kg. There is no specific freight classification.

Analytical Methods. Most analytical methods use the oxidizing power of iodine for its determination. The results are generally expressed as an equivalent concentration of elemental iodine. The choice of a method for the analysis of iodine depends on the concentration range to be determined.

Thiosulfate titration of iodine is limited to an iodine concentration of 7.5 µg mL (69). The use of organic solvents such as benzene, toluene, chloroform, and carbon tetrachloride as indicators in the titration of iodine have been proposed (70–72). These procedures increase the sensitivity of the titration so that 6.0 µg mL of iodine can be detected, although a sensitivity of 2 µg mL has been claimed (73).

Arsenious oxide, trivalent antimony (73), sulfurous acid (74), hydrogen sulfide (75), stannous ion, and thiocyanate (76) have been recommended for the titration of iodine. However, none of these appears to have a greater sensitivity for the determination of minute quantities of iodine than thiosulfate. Organic compounds such as formaldehyde (77), chloral hydrate (78), aldoses (79), acetone (70,80), and hydroquinone have also been suggested for this purpose.

Titration methods using adsorption indicators, based on the precipitation of insoluble iodides, have also been proposed (81–84). The sensitivity of these methods is less than that for the thiosulfate titration. Electrometric titration of the reaction between iodine and thiosulfate (85) was not found to be practicable for routine determinations of minute quantities of iodine.

The methods in which iodine is used as a catalyst for the reaction between ceric sulfate and nitrite or arsenite (86,87) are capable of determining small amounts of iodine. However, these catalytic methods are delicate and require accurate timing, careful temperature control, and special apparatus.

In view of the chromophoric character of the elemental iodine itself, many colorimetric methods have been proposed for the determination of inorganic iodine (88–92). These methods use the visible portion of the spectrum in reading iodine concentrations. In the visible range the extinction coefficient for iodine is not high enough to be used for minute quantities of iodine in water and other solvents (93). Higher sensitivities have been reported for elemental iodine in potassium iodide solutions in the ultraviolet (93,94).

Methods for iodine determination in foods using colorimetry (95,96), ion-selective electrodes (94,97), micro acid digestion methods (98), and gas chromatography (99) suffer some limitations such as potential interferences, possibility of contamination, and loss during analysis. More recently neutron activation analysis, which is probably the most sensitive analytical technique for determining iodine, has also been used (100–102).

Iodide ion, a moderately effective reducing agent, is used extensively for the determination of oxidants. In such applications, the iodine liberated by reaction between the analyte and the unmeasured excess of potassium iodide is ordinarily titrated with a standard solution of sodium thiosulfate. The reaction is as follows:



The quantitative conversion of thiosulfate to tetrathionate is unique with iodine. Other oxidant agents tend to carry the oxidation further to sulfate ion or to a mixture of tetrathionate and sulfate ions. Thiosulfate titration of iodine is best performed in neutral or slightly acidic solutions. If strongly acidic solutions must be titrated, air oxidation of the excess of iodide must be prevented by blanketing the solution with an inert gas, such as carbon dioxide or

nitrogen. One simple way of providing a blanket of carbon dioxide is to introduce a quantity of solid sodium carbonate into the solution, which reacts to form a layer of carbon dioxide that excludes oxygen from the titration vessel (103).

Health and Safety Factors and Regulations. Iodine is much safer to handle at ordinary temperatures than the other halogens because iodine is a solid and its vapor pressure is only 1 kPa (7.5 mm Hg) at 25°C, compared to 28.7 kPa (215 mm Hg) for bromine and 700 kPa (6.91 atm) for chlorine. When handling properly packed containers, usual work clothes are sufficient. In the handling of solid, unpacked iodine, rubber gloves, rubber apron, and safety goggles are recommended. Respirators or masks are also recommended.

The U.S. Occupational Safety and Health Administration (OSHA) has set a ceiling level for iodine of 0.1 ppm in air. The American Conference of Government and Industrial Hygienists (ACGIH) established 0.1 ppm as the TLV (TWA) for iodine. The maximum allowable concentration in air (MAK value) is also 0.1 ppm (104–106).

Empty containers may be destroyed in an incinerator or decontaminated by washing with a dilute thiosulfate or sulfite solution. Bulk wastes should be treated by controlled iodine recovery processes.

Iodine can affect the body if inhaled, if it comes in contact with the eyes or skin, or if it is swallowed. It may enter the body through the skin. Iodine vapor is a severe irritant of the eyes, respiratory tract, and to a lesser extent, to the skin. Swallowing iodine may cause burning in the mouth, vomiting, abdominal pain, and diarrhea. Short contact of iodine with the skin may produce a severe irritation of the skin and coloration similar to that obtained when tincture of iodine is applied to a wound. Prolonged contact can be harmful and may cause burns.

As an emergency treatment, the washing of the contaminated body parts with a 5% thiosulfate solution is recommended. If swallowed, gastric lavage with 5% solution of thiosulfate, followed by saline catharsis should be accomplished. If pulmonary signs are severe, oxygen should be supplied with intermittent positive-pressure breathing apparatus.

Chronic absorption of iodine causes iodism characterized by insomnia, inflammation of the eyes and nose, bronchitis, tremor, diarrhea, and weight loss.

Iodine is not combustible by itself, but can react very vigorously with reducing materials. It is incompatible with acetaldehyde, C₂H₂, Al, NH₃, NH₄, OH, Sb, BrF₃, CsHC₂, CsC₂, Cs₂O, CuC₂, ethanol, HgO, O₂, Li, LiC₂, Li C, Mg, F₂O, K, RbHC₂, Rb₂C₂, AgN₃, NaH, ZrC, SO₃, formamide, pyridine, P, ethanol and butadiene, and ethanol plus methanol and HgO.

Inorganic Iodine Compounds

Iodides. Iodides range from the completely ionic such as potassium iodide [7681-11-0], KI, to the covalent such as titanium tetraiodide [7720-83-4], TiI₄. Commercially, iodides are the most important class of iodine compounds. In general, these are very soluble in water and some are hygroscopic. However, some iodides such as the cuprous, lead, silver and mercurous, are insoluble.

The iodides have less tendency to form complexes than other halides, but complexes with mercury and several other elements, eg, cadmium and platinum, are known. Among the iodomercuriates are KHgI₃ [7783-33-7], the basis of Nessler's reagent for the detection of ammonia and Mayer's reagent for alkaloids (qv), Cu₂HgI₄ [13876-85-2], and Ag₂HgI₄ [7784-03-4]. These last two complexes undergo allotropic transformations at 70 and 50°C, respectively, with marked changes of color, and have been used as chromogenic materials. Silver iodomercuriate(II) [7784-03-4] has the highest electric conductivity at ordinary temperatures of any known solid.

The iodides of the alkali metals and those of the heavier alkaline earths are resistant to oxygen on heating, but most others can be roasted to oxide in air and oxygen. The vapors of the most volatile iodides, such as those of aluminum and titanium(II) actually burn in air. The iodides resemble the sulfides in this respect, with the important difference that the iodine is volatilized, not as an oxide, but as the free element, which can be recovered as such. Chlorine and bromine readily displace iodine from the iodides, converting them to the corresponding chlorides and bromides.

Alkali Metal Iodides. Potassium iodide [7681-11-0], KI, mol wt 166.02, mp 686°C, 76.45° I, forms colorless cubic crystals, which are soluble in water, ethanol, methanol, and acetone. KI is used in animal feeds, catalysts, photographic chemicals, for sanitation, and for radiation treatment of radiation poisoning resulting from nuclear accidents. Potassium iodide is prepared by reaction of potassium hydroxide and iodine, from HI and KHCO₃, or by electrolytic processes (107,108). The product is purified by crystallization from water (see also FEEDS AND FEED ADDITIVES; PHOTOGRAPHY).

Sodium iodide [7681-82-5], NaI, mol wt 149.92, mp 662°C, 84.66° I, forms colorless cubic crystals, which are soluble in water, ethanol, methanol, and acetone. It is used in photography, for the production of organic chemicals, and as an expectorant in cough medicines. NaI is separated by addition of sodium hydroxide or sodium carbonate to an acidic iodide solution (see also EXPECTORANTS, ANTITUSSIVES, AND RELATED AGENTS).

Hydrogen Iodide. Hydrogen iodide [10034-85-2], HI, mol wt 127.93, mp -50.9°C, bp -35.1°C, 99.21° I, is a colorless, nonflammable gas that fumes in moist air and is decomposed by light. It is unstable at room temperatures and above, slowly decomposing to hydrogen and iodine. It is extremely soluble in water forming an azeotrope: 234 g HI in 100 g H₂O at 10°C and 900 g HI in 100 g H₂O at 0°C. Hydrogen iodide is prepared by catalytic reaction of iodine and hydrogen, or concentrated HI solutions with P₂O₅ (109–112). Hydrogen iodide is used in the manufacture of hydroiodic acid, organic iodo compounds, and to remove the iodine from iodo compounds.

Hydroiodic acid, the colorless solution formed when hydrogen iodide gas dissolves in water, is prepared by reaction of iodine with hydrogen sulfide or hydrazine; or by an electrolytic method. Typically commercial hydroiodic acid contains 40–55% HI. Hydroiodic acid is used in the preparation of iodides and many organic iodo compounds.

Iodates. Iodates are stable at room temperatures but lose oxygen on heating. Metallic iodates, although stable and safe to handle, should be kept out of contact with organic substances and other combustible materials, because such mixtures are explosive. Iodates can be prepared by oxidation of iodine to iodic acid, followed by neutralization with an oxide or hydroxide; or by electrolytic oxidation of an iodide solution (105,106).

Potassium iodate [7758-05-6], KIO₃, mol wt 214.02, 59.30° I, forms white, odorless crystals or a crystalline powder. It has a density 3.98 g/mL and mp of 560°C with partial decomposition. Potassium iodate is rapidly formed when potassium iodide is fused with potassium chlorate, bromate, or perchlorate. The solubility in water is 9.16 g/100 g H₂O at 25°C and 32.2 g/100 g H₂O at 100°C. KIO₃ is extensively used as an oxidizing agent in analytical chemistry; and as a maturing agent and dough conditioner (see BAKERY PROCESSES AND LEAVENING AGENTS).

Sodium iodate [7681-55-2], NaIO₃, mol wt 197.90, 64.13° I, is a white crystalline powder. Its solubility in water is similar to that of potassium iodate: 9.0 g/100 g H₂O at 25°C and 34.0 g/100 g H₂O at 100°C. It is insoluble in alcohol. Two hydrates exist in saturated solution, the pentahydrate up to 19.85°C, and the monohydrate from 19.85 to 73.4°C.

Calcium iodate monohydrate [10031-32-0], Ca(IO₃)₂·H₂O, mol wt 407.90, 62.22° I, is a solid white powder having a density of about 4.5 g/mL. The monohydrate is stable up to 540°C, but it is very sensitive to reducing agents. It is slightly soluble in water, insoluble in alcohol, and more soluble in aqueous solutions of iodides. It is mainly used in animal and fowl feeds.

Iodic acid [7782-68-5], HIO₃, mol wt 179.93, 72.14° I, mp 110°C (dec), density 4.65 g/mL, forms orthorhombic crystals that darken upon exposure to light, but it is not hygroscopic. It is soluble in nitric acid and diluted alcohol and very soluble in water: 269 g/100 mL H₂O at 20°C and 295 g/100 mL H₂O at 40°C. It is insoluble in absolute alcohol, ether, and chloroform. HIO₃ decomposes to HIO₃·I₂O₅ at 70°C, and decomposes completely to I₂O₅ at 200°C. HIO₃ is prepared by oxidation of iodine with perchloric acid, nitric acid, or hydrogen peroxide; or oxidation of iodine in aqueous suspension to iodic acid by silver nitrate. Iodic acid is also formed by anodic oxidation at a platinum electrode of iodine dissolved in hydrochloric acid (113,114).

Iodine Halides and Polyhalides. Iodine forms six well-defined compounds with the other halides (115,116). These compounds are readily formed by direct reaction of the two halogens (117).

Polyhalide compounds are formed between iodine or iodine halides and other halide salts. The formulas of some of these compounds are RbI₃ [12298-69-0], KI₃·H₂O [7790-42-3], KICl₄ [14323-44-5], KIBrCl [15859-97-9], N(CH₃)₄I₉ [3345-37-7], and N(C₂H₅)₄IBr₂ [20445-98-1]. These are crystalline

compounds, usually orthorhombic, ranging in color from deep black through red and orange to yellow and even white.

Iodine monochloride *[7790-99-0]*, ICl, mol wt 162.38, 78.16° o I, is a black crystalline solid or a reddish brown liquid. Solid ICl exists in two crystalline modifications: the α-form, as stable ruby-red needles, *d* = 3.86 g/mL and mp 27.3°C; and as metastable brownish red platelets, *d* = 3.66 g/mL, mp 13.9°C and bp 100°C (dec). Iodine monochloride is used as a halogenation catalyst and as an analytical reagent (Wij’s solution) to determine iodine values of fats and oils (see FATS AND FATTY OILS). ICl is prepared by direct reaction of iodine and liquid chlorine. Aqueous solutions are obtained by treating a suspension of iodine in moderately strong hydrochloric acid with chlorine gas or iodic acid (118,119).

Iodine trichloride *[865-44-1]*, ICl₃, mol wt 233.39, 54.40° o I, is a yellow or brownish powder. It is pungent and has a very irritating odor. It decomposes at 77°C into ICl and Cl₂. It is prepared by adding finely powdered iodine to an excess of liquid chlorine. It is used as a chlorinating and oxidizing agent (120).

Other iodine halides are iodine monobromide *[7789-33-5]*, IBr, iodine tribromide *[7789-58-4]*, IBr₃, iodine pentafluoride *[7783-66-6]*, IF₅, and iodine heptafluoride *[16921-96-3]*, IF₇.

Organic Iodine Compounds

The organic iodine compounds have lower heats of formation and greater reactivities than their chloro and bromo analogues. As in the case of the inorganic iodides, their indexes of refraction and specific gravities are higher than the corresponding chloro and bromo derivatives (121).

The aliphatic iodine derivatives are usually prepared by reaction of an alcohol with hydroiodic acid or phosphorus triiodide; by reaction of iodine, an alcohol, and red phosphorus; addition of iodine monochloride, monobromide, or iodine to an olefin; replacement reaction by heating the chlorine or bromine compound with an alkali iodide in a suitable solvent; and the reaction of triphenyl phosphite with methyl iodide and an alcohol. The aromatic iodine derivatives are prepared by reacting iodine and the aromatic system with oxidizing agents such as nitric acid, fuming sulfuric acid, or mercuric oxide.

Methane Derivatives. Methyl iodide *[74-88-4]*, also known as iodomethane, CH₃I, mol wt 141.95, 89.41° o I, is a colorless, pungent liquid having a density of 2.279 g/mL at 20°C, a melting point of -66.1°C, and a boiling point of 42.5°C, that turns brown on exposure to light. It is poisonous. Its solubility in water is 1.4 g/100 g H₂O at 20°C, and it is miscible with alcohol and ether. Methyl iodide is prepared by reaction of methanol with phosphorus and iodine; from potassium iodide and methyl sulfate or methyl *p*-toluenesulfonate; by reaction of dimethyl sulfate with an aqueous iodine slurry containing a reducing agent such as iron or sodium bisulfite; by reaction of methanol and hydrogen iodide; and by reaction of methanol, iodine, and diborane (122,123). It is used as a methylation agent in organic synthesis, in microscopy (qv) because of its high refractive index, as an embedding material for examining diatoms, and in testing for pyridine.

Methylene iodide *[75-11-6]*, CH₂I₂, also known as diiodomethane, mol wt 267.87, 94.76° o I, mp 6.0°C, and bp 181°C, is a very heavy colorless liquid. It has a density of 3.325 g/mL at 20°C and a refractive index of 1.7538 at 4°C. It darkens in contact with air, moisture, and light. Its solubility in water is 1.42 g/100 g H₂O at 20°C; it is soluble in alcohol, chloroform, benzene, and ether. Methylene iodide is prepared by reaction of sodium arsenite and iodoform with sodium hydroxide; reaction of iodine, sodium ethoxide, and hydroiodic acid on iodoform; the oxidation of iodoacetic acid with potassium persulfate; and by reaction of potassium iodide and methylene chloride (124,125). Diiodoform is used for determining the density and refractive index of minerals. It is also used as a starting material in the manufacture of x-ray contrast media and other synthetic pharmaceuticals (qv).

Iodoform *[75-47-8]*, CHI₃, is also known as triiodomethane, mol wt 393.78, 96.69° o I, mp about 120°C, and 4.008 g/mL density at 20°C. Iodoform is a yellow crystalline powder having an unctuous touch and disagreeable odor that decomposes at high temperatures, with iodine evolution. It is only slightly soluble in water: 0.01 g/100 g H₂O at 20°C. It is soluble in chloroform, benzene, and glycerol. Iodoform is prepared by reaction of acetone, sodium hypochlorite, iodine, and sodium hydroxide; by reacting chloroform and methyl iodide; and by electrolysis of an iodide solution in dilute alcohol or acetone (126). Iodoform is used as wound dressing and as a sensitizing agent in certain printing processes (qv).

Other Iodo Compounds. Thymol iodide, C₂OH₂₄I₂O₂, mol wt 550.23, 46.13° o I, is a reddish brown or reddish yellow bulky powder that gives off iodine vapors when heated above 100°C and upon prolonged exposure to light. Readily soluble in chloroform, ether, collodion, and volatile oils, it is slightly soluble in alcohol and insoluble in water, glycerol, and alkaline solutions. Thymol iodide is prepared by treating a solution of thymol in sodium hydroxide with a potassium iodide–iodine solution. Its principal use is as an antiseptic disinfectant (127–129).

Ethyl iodide *[75-03-6]*, C₂H₅I, also known as iodoethane, is a colorless liquid having a density of 1.933 g/mL at 20°C and a boiling point of 72.2°C. Because of its high density, it is used in petrology as a heavy liquid for determining the density of rock and mineral fragments. In medicine it has been used for the treatment of fungous diseases of the skin. It is only slightly soluble in water, but completely soluble in alcohol and ether. Ethyl iodide is prepared by reactions similar to those used for methyl iodide (122).

Iodobenzene *[591-50-4]*, C₆H₅I, mol wt 204.02, 62.23° o I, mp -30°C, bp 188–189°C, is a colorless liquid that rapidly becomes yellow and has a characteristic odor. It is insoluble in water, but completely miscible with alcohol, chloroform, and ether. It has a density of 1.832 g/mL at 20°C and a refractive index of 1.621 at 4°C. Iodobenzene is prepared by the reaction of iodine and benzene in the presence of an oxidizing agent; and from benzenediazonium sulfate and potassium iodide (122). Iodobenzene is used as a heavy liquid for refractive index determinations, but probably its principal use is in the synthesis of iodoso compounds, RIO; iodoxy compounds, RIO₂; and iodonium salts, R₂IX.

Iodophors (tamed iodine) are compounds in which surface-active agents, such as nonoxynol, act as carriers and solubilizing agents for iodine. Iodophors usually enhance the bactericidal activity of iodine and reduce its vapor pressure and odor. Further, staining is avoided and high dilution with water is possible (130–132).

Uses

Iodine has a wide range of uses in the chemical and allied industries. A high percentage of the initial use of iodine lies in the production of intermediates, which are frequently marketed as such. A breakdown of world iodine consumption for 1992 suggests that 35° o was used to manufacture pharmaceuticals, 30° o for inorganic salts, 15° o as sanitizers (iodophors), 12° o was used to produce other organic derivatives, 5° o for agricultural chemicals, and 3° o for catalysts. The expected annual increase in iodine consumption is about 5° o in the area of pharmaceuticals and sanitizers, and 2° o for the inorganic salts. An increase of 1–2° o is estimated for the other areas.

Industrial.
Photography. Photography (qv) represents one of the oldest industrial uses of iodide. The sensitive silver salt in rapid negative emulsions contains up to 7° o or more silver iodide *[7783-96-2]*, AgI. From 1969 to 1985 estimates on iodine consumption for this purpose varied from 150 to 270 t/yr (66). Triphenylphosphonium iodide is also among the iodine derivatives used in photography. This derivative permits faster development and higher contrast photography.

Dyes, Inks, and Colorants. Some dyes contain iodine, including 4',5'-diiodoflourescein *[38577-97-8]*, rose bengal *[11121-48-5]*, and erythrosin *[16423-68-0]* (see XANTHENE DYES), and members of the cyanine group (see CYANINE DYES). Erythrosine FD&C Red No. 3 (disodium salt of 9-*O*-carboxiphenyl-6-hydroxy-2',4',5',7'-tetraiodo-3'-isoxanthrone) is an orthochromatic sensitizer for photographic emulsions and also a certified food colorant (see COLORANTS FOR FOODS, DRUGS, COSMETICS, AND MEDICAL DEVICES). Iodine consumption in erythrosine was estimated to be ca 116 t in 1987 (66).

Catalysts. Iodine and its compounds are very active catalysts for many reactions (133). The principal use is in the production of synthetic rubber via Ziegler-Natta catalysts systems. Also, iodine and certain iodides, eg, titanium tetraiodide *[7720-83-4]*, are employed for producing stereospecific polymers, such as polybutadiene rubber (134); about 75° o of the iodine consumed in catalysts is assumed to be used for polybutadiene and polyisoprene polymerization (66) (see RUBBER CHEMICALS). Hydrogen iodide is used as a catalyst in the manufacture of acetic acid from methanol (66). A 99° o yield as acetic acid has been reported. In the heat stabilization of nylon suitable for tire cordage, iodine is used in a system involving copper acetate or borate, and potassium iodide (66) (see TIRE CORDS).

When heated with small amounts of iodine, rosins, tall oil, and other wood products are converted to more stable forms (135,136). Iodine has been used with some tin salts as a catalyst in the hydrogenation of coal (qv) and its distillation products (137,138), and has been recommended as a catalyst for the production of drying oils (qv) from unsaturated animal fats (139,140).

Alkali metal and other iodides are effective catalysts in reactions involving aliphatic chloro and bromo compounds, such as the preparation of cyclopropane from 1,3-dichloropropane and metallic zinc (141).

Iodine catalyzes the conversion of amorphous selenium to the black, semiconducting metallic modification, and is used for this purpose in the manufacture of photoelectric cells and electric rectifiers (see SELENIUM AND SELENIUM COMPOUNDS).

Minor industrial uses include the application of silver iodide as a smoke for the seeding of clouds to induce rainfall. Compounds used for obtaining some nonflammable plastics and cellulose are benzyltriphenyl-phosphonium iodides and [2'-(acetyloxy)ethyl] triphenyl-phosphonium iodides (see FLAME RETARDANTS, HALOGENATED FLAME RETARDANTS) (142). The addition of iodine to an aromatic hydrocarbon such as *n*-butylbenzene results in the formation of charge-transfer complexes that display outstanding effectiveness as lubricants for hard-to-lubricate metals (143), such as titanium or steels (see also LUBRICATION AND LUBRICANTS). Iodine is also used in the production of high purity metals such as titanium, silicon, hafnium, and zirconium (144).

Nutrition. Because of the requirements of the thyroid gland for the synthesis of the hormone thyroxine [51-48-9], iodine is an essential nutrient (see MINERAL NUTRIENTS). Endemic iodine deficiency results in the condition known as goiter. To ensure uptake iodine is added to table salt for human consumption and to animal feedstuffs. To prevent goiter and to increase yields of milk from dairy cattle and eggs from fowl, iodized protein is added to animal feeds. Supplements of iodine are particularly important in breeding cows and calves. Iodine deficiency during pregnancy and lactation increases the incidence of weak or dead births in calves, pigs, horses, and kids. The four iodine derivatives accounting for nearly all nutrient iodine consumption are potassium iodide, calcium iodate [7789-80-2], ethylenediamine dihydriodide [5700-49-2] (EDDI), and iodophores (iodine complexes) (see FOOD ADDITIVES). World consumption of iodine in animal feedstuffs has been estimated to be between 8° (145) and 13° (146) of iodine usage, the majority of it being used in the United States and Western Europe.

It is difficult to define the normal range of iodine intake in humans, and despite efforts to provide iodine supplementation in many geographic areas of the world, endemic iodine deficiency and attendant goiter remain a world health problem (147). Exposure to excess iodine may sometimes lead to the development of thyroid disease. This unusual type of iodide-induced goiter has been found, for example, in 10° of the population of a Japanese island where fishermen and their families consume large quantities of an iodine-rich seaweed and have an iodine intake as high as 200 mg d (148).

Pharmaceuticals.

Medicinals and Sanitation. Iodine and iodine compounds and preparations are employed extensively in medicine, eg, as antiseptics, as drugs administered in different combinations in the prophylaxis and treatment of diseases, and as therapeutic agents in various thyroid dyscarsias and other abnormalities. The principal active ingredients containing iodine have been classified as antiseptics, antispasmodics, coronary vasodilators, diagnostics, endocrinology, and neuromuscular blocking agents, and are used in the fields of gastroenterology, metabolism and nutrition, neurology–psychiatry, ophthalmology, parasitology, pneumology, and rheumatology (143).

To prevent radioactive iodides from lodging in the thyroid gland during exposure to excessive radiation, a potential application of iodine acting as a thyroid-blocker has arisen. For this purpose potassium iodide was recommended (66).

Iodine is extensively used in a variety of forms as both an antiseptic and a disinfectant. Iodophors, usually nonionic surfactants (qv) complexed with iodine, were developed for more readily usable iodine-based antiseptics and disinfectants. These are used as disinfectants in dairies, laboratories, and food processing (qv) plants, and for sanitation of dishes in restaurants. The reaction product of lanolin and iodine shows utility as a germicide (149).

A poly(*n*-vinyl-2-pyrrolidinone)-iodine complex [25655-41-8] (PVP-iodine), has been used extensively in hospitals and elsewhere because of its germicidal, bactericidal, fungicidal, and generally disinfecting properties (150). It is sold as a solution that contains about 10° available, or active, iodine and about 5° inactive iodine, in the form of iodide ion (see DISINFECTANTS AND ANTISEPTICS; INDUSTRIAL ANTIMICROBIAL AGENTS).

Radioactive Isotopes. Iodine radioactive isotopes emit beta- or gamma-radiation (66). Radioactive iodine has been used successfully for the treatment of cancer of the thyroid (see RADIOPHARMACEUTICALS). Iodine-131 [10043-66-0] is the most widely available and highly used isotope for the diagnosis and therapy of thyroid disorders (see RADIOACTIVE TRACERS). Iodine-125 [14158-31-7] has been useful for the delineation of superficial lesions. Iodine-123 [15715-08-9], which is produced in a cyclotron, is considered the agent of choice for imaging and uptake measurements despite its high cost, because it minimizes the patient's radioactive exposure; its half-life time is only 13 h and it emits a 159-keV photon (147) (see MEDICAL IMAGING TECHNOLOGY).

Water Purification. Iodine effectively disinfects water against bacteria, viruses, and cysts. Globaline tablets were developed for the disinfection of small or individual water supplies in the U.S. Army during World War II. Studies of the disinfection of public water supplies using iodine were initiated in 1963 (151). These studies showed that concentrations up to 5 ppm of iodine were not deleterious to health and that 1 ppm was sufficient to safely disinfect a water supply at the high pH values encountered in many treated supplies, with virtually no loss of effectiveness because of iodate buildup. Administration of approximately 2 mg of iodide in a single dosage induces the Wolff-Chaikoff effect (152), ie, the inhibitory effect of iodine on hormone formation by the thyroid. Commercially available iodinators control potentially dangerous organisms by passing a side stream of water through a bed of prilled iodine to provide 0.5 ppm iodine in a water supply.

Iodine may also be used as an effective microbicide for swimming pools (151,153,154). It acts like free chlorine and is superior to chloroamines in bacterial, virucidal, and cysticidal efficiency (see also WATER, TREATMENT OF SWIMMING POOLS, SPAS, AND HOT TUBS). In addition, iodine shows decreased reactivity with organic matter, less variation in microbicidal efficiency over the typical pH range observed in pools, and potential for regeneration of free iodine through application of a suitable oxidant, eg, chlorine, ozone, or potassium monopersulfate. Other advantages of iodine over chlorine are its longer life, lack of odor, no bleaching action, and small amount of eye irritation. Unfortunately, dependable control of algae proliferation in well-illuminated swimming pools has not been accomplished with iodine alone. Controlled field experiments have shown that compounds such as prometryne and terbutryne are effective algicides compatible (nonreactive) with free iodine as microbicide (155). Chlorine can be used in combination with iodine as a means of controlling algae and to reoxidize residual iodide to free iodine.

Agricultural.

Herbicides. The use of herbicides (qv) based on iodine compounds has its main market in Western Europe. In Canada and the United States these compounds are used only to a small extent. The only significant iodine-containing herbicide is ioxynil [1689-83-4] (4'-hydroxy-3',5'-diiodobenzoic acid). This compound, often used in combination with other herbicides, is formulated for controlling many annual broad-leaved weeds, especially black-bindweed, knotgrass, mayweeds, and corn marigold post-emergence in wheat, barley, oats, rye, and triticale (142). Annual consumption of iodine in relation to ioxynils is considered to be about 300–500 t (66).

Insecticides. The use of iodine-based compounds as insecticides is of minor importance. The active ingredient is Iodofenphos (142). It is formulated to be used in public health and animal husbandry, eg, for cockroach control (see INSECT CONTROL TECHNOLOGY).

BIBLIOGRAPHY

"Iodine and Iodine Compounds" in *ECT* 1st ed., Vol. 7, pp. 942–972, by A. C. Loonam, Deutsch and Loonam, for Chilean Iodine Educational Bureau, Inc., and M. G. Gergel and M. Revelise, Columbia Organic Chemicals Co., Inc.; "Iodine Preparations" in *ECT* 1st ed., Vol. 7, pp. 972–981, by L. Gershenfeld, Philadelphia College of Pharmacy and Science; "Iodine and Iodine Compounds" in *ECT* 2nd ed., Vol. 11, pp. 847–870, by A. W. Hart, The Dow Chemical Co., and M. G. Gergel and J. Clark, Columbia Organic Chemicals Co., University of South Carolina; in *ECT* 3rd ed., Vol. 13, pp. 649–677, by C. J. Mazac, PPG Industries.

1. J. L. Gay-Lussac, *Ann. Chim.* **91**, 5 (1814).
2. M. Van Mons, *Ann. Phil.* **3**, 429 (1814).
3. *Geochemistry of Iodine*, Chilean Iodine Educational Bureau, The Shenval Press, London, 1956.
4. G. Lunde, *Beitr. Geophys.* **19**, 99 (1928).

5. *Iodine Facts Nos. 235–237, Nos. 244–246*, Chilean Iodine Educational Bureau, London, May 1943.
6. J. F. McClendon, *Iodine and the Incidence of Goitre*, University of Minnesota Press, Minneapolis, 1939.
7. E. S. Grew, *Inorg. Chem.* **28**, 23 (1946).
8. *Iodine Facts Nos. 229–233*, Chilean Iodine Educational Bureau, London, May 1943.
9. M. Carranza, *Bol. Soc. Quim. Perù* **2**, 199 (1936).
10. G. Lunde, *Beitr. Geophys.* **16**, 413 (1927).
11. G. E. Ericksen, *Geology of the Salt Deposits and the Salt Industry of the Northern Chile*, U.N. Special Fund Chile Mineral Survey Project, Washington, D.C., Apr. 1963.
12. H. Hansen, *Ann. Phys.* **35**, 524 (1939).
13. W. Meyer, *Ann. Phys.* **36**, 651 (1939).
14. M. Starumanis and J. Sanka, *Z. Phys. Chem.* **B53**, 320 (1943).
15. W. Theilacker, *Z. Naturforsch.* **3B**, 233 (1948).
16. T. Nayder, *Bull. Int. Acad. Polon.* **A**, 231 (1934).
17. A. Sapper and W. Biltz, *Z. Anorg. Allg. Chem.* **198**, 184 (1931).
18. H. Braune and R. Linke, *Z. Anorg. Allg. Chem.* **148A**, 195 (1930).
19. H. Braune and R. Linke, *Z. Phys. Chem.* **148A**, 242 (1940).
20. H. Brown, *Phys. Rev.* **57**, 242 (1940).
21. M. Starumanis and J. Sanka, *Z. Phys. Chem.* **B53**, 320 (1943).
22. G. Careti, *Nuovo Cimento* **6**, 94 (1949).
23. H. Razow, *Z. Anorg. Chem.* **114**, 117 (1920).
24. H. B. Baker, *J. Chem. Soc.* **121**, 568 (1922).
25. E. Hendrick, *Chem. Met. Eng.* **27**, 544 (1922).
26. W. Guertler and M. Pirani, *Z. Metallk.* **11**, 1 (1919).
27. H. T. Guery and L. Gillespie, *J. Phys. Rev.* **40**, 269 (1932).
28. L. Gillespie and L. H. D. Fraser, *J. Amer. Chem. Soc.* **58**, 2260 (1936).
29. B. Topley, *Nature* **118**, 802 (1926).
30. F. Lange, *Z. Phys. Chem.* **110**, 343 (1924).
31. K. J. Frederick and J. H. Hildebrand, *J. Amer. Chem. Soc.* **60**, 1436 (1938).
32. L. G. Carpenter and T. F. Hade, *Phil. Mag.* **23**, 193 (1923).
33. A. Pochettino, G. Fulcheris, and R. Atti, *Acad. Torino* **58**, 493 (1923).
34. T. J. Webb, *J. Phys. Chem.* **29**, 816 (1925).
35. K. K. Kelly, *Bur. Mines Bull.* **350**, 63 (1932).
36. O. B. E. Stephens and T. Stephens, *The Solubilities of Inorganic and Organic Compounds*, Macmillan Co., New York, 1964.
37. E. Roth and G. Schay, *Z. Phys. Chem.* **B28**, 323 (1935).
38. H. Otuka and G. Schlay, *Z. Phys. Chem.* **B1**, 62 (1928).
39. E. Montignie, *Bull. Soc. Chim.* **7** **8**, 747 (1947).
40. F. Korosy, *J. Amer. Chem. Soc.* **61**, 838 (1939).
41. E. I. Gurovic, *Bull. Acad. Sci. USSR*, 1453 (1937).
42. A. W. Spitz, *Chem. Eng.* **54**, 135 (1947).
43. Y. Klinov, *Khim. Prom.* **10/11**, 14 (1944).
44. M. Dode and J. Basset, *Bull. Soc. Chim.* **2**, 344 (1935).
45. W. M. MacKevin and W. M. Carson, *J. Amer. Chem. Soc.* **72**, 42 (1950).
46. H. Pelabon, *Compt. Rend.* **189**, 1085 (1929).
47. C. R. DeRobles and E. Moles, *An. Fis. Quim.* **32**, 474 (1934).
48. D. A. Palmer and M. H. Lietzke, *Radiochim. Acta* **31**, 37 (1982).
49. I. Lengyel, I. R. Epstein, and K. Kustin, *Inorg. Chem.* **32**, 5880–5882 (1993).
50. W. M. Latimer, *The Oxidation Potentials of the Elements and their Potentials in Aqueous Solutions*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1938.
51. G. Jones and B. B. Kaplan, *J. Amer. Chem. Soc.* **50**, 2066 (1928).
52. J. C. Bailar, H. J. Emelens, R. Nyhom, and A. F. Trotman-Dickenson, eds., *Comprehensive Inorganic Chemistry*, Pergamon Press, Elmsford, N.Y., 1973.
53. E. Abel and F. Stadler, *Z. Phys. Chem.* **122**, 49 (1926).
54. E. Abel and K. Hilferding, *Z. Phys. Chem.* **136**, 186 (1928).
55. P. Putzeys, *Ann. Soc. Sci. Bruxelles* **47B**, 159 (1927).
56. F. Bush, *J. Phys. Chem.* **33**, 613 (1929).
57. P. A. Lynday, *Ind. Minerals*, Mar. 1986.
58. G. E. Ericksen, *Geology and Origin of the Chilean Nitrate Deposits*, U.S. Government Printing Office, Washington, D.C., 1981.
59. D. A. Garret, *Chemistry and Origin of the Chilean Nitrate Deposits*, Sixth International Symposium on Salt, Vol. I, 1981.
60. E. Mucoz, *SQM Iodine Internal Report*, SQM Iodine, 1992.
61. L. G. Gindin and M. V. Pavlova, *Dokl. Akad. Nauk. SSSR* **69**, 377 (1949).
62. L. Ya. Klinov, *Khim. Prom.* **10/11**, 17 (1944).
63. V. I. Ginzburg and O. I. Kabakova, *Zashch. Met.* **5**, 627 (1969).
64. F. G. Sawyer, F. G. Ohman, and F. E. Lush, *Ind. Eng. Chem.* **41**, 1547 (1949).
65. *Chem. Mkt. Rep.* Dec. 1990.
66. *The Economics of Iodine*, 5th ed., Roskill Information Services Ltd., London, Feb. 1989.
67. *American Chemical Society Specifications, Reagent Chemicals*, 7th ed., American Chemical Society, Washington, D.C., 1986.
68. USP XXII, *The United States Pharmacopeia*, 22nd rev., United States Pharmacopeial Convention, Inc., Rockville, Md., Jan. 1990.
69. I. M. Kolthoff, H. Menzel, and N. H. Furman, *Volumetric Analysis*, Vol. II, John Wiley & Sons, Inc., New York, 1929.
70. H. Dietz and B. M. Margoshes, *Chem. Ztg.* **28**, 1191 (1904).
71. Rebourdin, *Compt. Rend.* **31**, 784 (1850); J. J. Custer and S. Natelson, *Anal. Chem.* **21**(4), 1005 (1949).
72. Schezow, *Z. Anorg. Chem.* **44**, 86 (1905).
73. I. M. Kolthoff and E. B. Sandell, *Textbook in Quantitative Inorganic Analysis*, Vol. II, Macmillan Co., New York, 1943.
74. A. Dupasquier, *Ann. Chim. Phys.* **73**, 310 (1840).
75. I. M. Kolthoff, *Z. Anal. Chem.* **60**, 451 (1921).
76. B. Kohler, *Chem. Listy* **14**, 137 (1920).
77. Romijn, *Z. Anal. Chem.* **39**, 60 (1900).
78. E. Rap, *Arch. Pharmacol.* **241**, 328 (1903).
79. Bland and Lloyd, *J. Soc. Chem. Ind.* **32**, 948 (1914).
80. Schezow, *Z. Anorg. Chem.* **44**, 86 (1905).

81. K. Fajans and O. Hassel, *Z. Electrochem.* **29**, 495 (1923).
82. K. Fajans and H. Wolff, *Z. Anorg. Allgem. Chem.* **137**, 221(1924).
83. F. Mohr, *Lehrbuch der Chemische Analytische Titrierenmetoden*, Vol. 2, S. Vieveg und Sohn, Braunschweig, Germany, 1859.
84. E. Vortecsek, *Chem. Ztg.* **42**, 257 (1918).
85. J. F. McClendon, *Iodine and the Incidence of Goitre*, University of Minnesota Press, 1939.
86. D. T. Englis and A. A. Knopfenmacher, *Ind. Eng. Chem. Anal. Ed.* **17**, 393 (1945).
87. H. S. Strichler and E. W. Strickler, *Endocrinology* **37**, 220 (1945).
88. W. N. Aldridge, *Analyst* **70**, 474 (1945).
89. J. F. Lox, I. Pitesky and A. S. Alving, *J. Biol. Chem.* **142**, 147 (1942).
90. Muller and Diffenthaler, *Z. Anorg. Allgem. Chem.* **67**, 418 (1910).
91. A. Saifer and J. Hughes, *J. Biol. Chem.* **118**, 214 (1937).
92. J. J. Custer and S. Natelson, *Anal. Chem.* **21**, 1005 (1949).
93. J. Sendroy and A. S. Alving, *J. Biol. Chem.* **142**, 159 (1942).
94. W. Brode, *J. Amer. Chem. Soc.* **48**, 1877 (1926).
95. P. Miles, *J. Assoc. Off. Anal. Chem.* **61**, 1366 (1978).
96. E. Sandell and I. M. Kolthoff, *Microchim. Acta* **1**, 9 (1973).
97. D. E. Lacroix, presented at *The Iodine in American Foods Workshop*, Scottsdale, Ariz., Nov. 13–14, 1979.
98. P. W. Fischer and M. R. Lr. Abbe, *J. Assoc. Off. Anal. Chem.* **64**, 71 (1931).
99. H. J. Baker, *J. Assoc. Off. Anal. Chem.* **60**, 1302 (1977).
100. O. Johansen and E. Seinnes, *Analyst* **101**, 455 (1976).
101. H. L. Rook, *J. Radioanal. Chem.* **39**, 351 (1977).
102. M. Allegrini, K. W. Boyer, and J. T. Tanner, *J. Assoc. Off. Anal. Chem.* **64**, 1111 (1981).
103. D. A. Skoog and D. M. West, *Analytical Chemistry: An Introduction*, Saunder College Publishing, New York, 1986.
104. *Occupational Health Guideline for Iodine*, (PB83-154609) U.S. Department of Health and Human Services, Washington, D.C., Sept. 1978.
105. H. E. Christensen and T. L. Luginbyhl, eds., *NIOSH Toxic Substances List*, 1974 Ed., Publication No. 74-134, Washington D.C., 1974.
106. *Iodine, Hygienic Guide Services*, American Industrial Hygiene Association, Detroit, Mich., 1965.
107. Lingane and I. M. Kolthoff, *Inorg. Syn.* **1**, 163 (1939).
108. **U.S. Pat. 2,989,450** (June 20, 1961), Marylott and Elkins.
109. Caley and Bulford, *Inorg. Syn.* **1**, 159 (1939).
110. Powell and Campbell, *J. Amer. Chem. Soc.* **69**, 1227 (1947).
111. G. Brauer, ed., *Handbook of Inorganic Preparative Inorganic Chemistry*, Academic Press, Inc., New York, 1963.
112. J. C. Bailar, H. J. Emelens, R. Nyhom, and A. F. Trotman-Dickenson, eds., *Comprehensive Inorganic Chemistry*, Pergamon Press, Elmsford, N.Y., 1973.
113. E. Moles and A. Parts, *An. Fis. Quim.* **31**, 618 (1933).
114. H. Willard and H. Ralston, *Trans. Amer. Electrochem. Soc.* **62**, 239 (1932); J. S. Carter and C. R. Hoskins, *J. Chem. Soc.*, 580 (1929).
115. S. I. Raikhshtein, L. Eventov, and I. A. Kazarnovskii, *Z. Anorg. Chem.* **216**, 1 (1933).
116. R. E. Gibson, *J. Chem. Phys.* **135**, 690 (1931).
117. Kornog and Karger, *Inorg. Syn.* **1**, 165 (1939).
118. Buckles and Bader, *Inorg. Syn.* **9**, 130 (1967).
119. Booth and Morris, *Inorg. Syn.* **1**, 167 (1939).
120. D. F. Banks, *Chem. Rev.* **66**, 243 (1966).
121. H. S. King, in A. H. Blatt, ed., *Organic Syntheses*, Collective Vol. II, John Wiley & Sons, Inc., New York, 1943.
122. N. Nagasi, *J. Pharm. Soc. Jpn.* **407**, 1 (1916).
123. Adams and Marvel, *Org. Syn.* **1**, 57 (1921).
124. Bagnara, *Eng. Mining J. Press* **116**, 51 (1923).
125. **Ger. Pat. 49,739**, Messinger and Vortmann; S. Budavari, ed., *The Merck Index*, Merck & Co., Inc. Rahway, N.J., 1989.
126. S. Glasstone, *Ind. Chem.* **7**, 315 (1931).
127. **Ger. Pat. 52,828**, Messinger and Vortmann; Ref. 125.
128. **Ger. Pat. 52,833**, Messinger and Vortmann; Ref. 125.
129. E. P. Blanchard and H. E. Simmons, *J. Amer. Chem. Soc.* **86**, 1337 (1964).
130. **U.S. Pat. 2,977,315** (Mar. 28, 1961), Schreib (to Lazarus Labs).
131. *Antara Chemicals*, BIOPAL VRO-20 Brochure, General Aniline and Film Corp., New York, 1958.
132. *Iodine Information, Catalytic Chemistry*, No. 31 and No. 32, Chilean Iodine Educational Bureau, London, 1957.
133. **Brit. Pat. 920,344** (Mar. 6, 1963), (to Phillips Petroleum Co.).
134. **Brit Pat. 931,440** (July 17, 1963), (to Phillips Petroleum Co.).
135. **U.S. Pat. 2,299,577** (Oct. 20, 1942), T. Hasselstrom and E. A. Brehnan (to G and A Labs. Inc.).
136. **U.S. Pat. 2,311,386** (Feb. 16, 1943), T. Hasselstrom (to G and A Labs Inc.).
137. K. Gordon, *Chem Age (London)* **55**, 761, 795 (1949).
138. **U.S. Pat. 2,452,271** (Mar. 15, 1949), H. R. Storch and L. I. Hirst.
139. **U.S. Pats. 2,411,111–2,411,113** (Nov. 12, 1946), A. W. Ralston and O. Turinsky (to Armour and Co.).
140. **U.S. Pat. 2,498,133** (Feb. 1, 1950), A. W. Ralston, O. Turinsky, and L. Van Akkeren (to Armour and Co.).
141. H. B. Haas, *Ind. Eng. Chem.* **28**, 1178 (1936).
142. *Etude bibliographique sur l'Iode et les Derives de l'Iode*, SQM Iodine Internal Document, SQM Iodine Corp., Norfolk, Va., Apr. 1993.
143. **U.S. Pat. 3,228,880** (Jan. 11, 1966) R. S. Owens and R. W. Roberts (to General Electric Co.).
144. R. F. Rolsten, *Iodide Metals and Metal Iodides*, John Wiley & Sons, Inc., New York, 1961.
145. P. A. Lyday, *Ind. Minerals*, April 1982.
146. *Chemical Week*, Jan. 1987.
147. S. C. Werner and S. H. Ingbar, eds., *The Thyroid, A Fundamental and Clinical Text*, Harper and Row, Hagerstown, Md., 1978.
148. *Brit. Med. J.*, 1566 (1977).
149. **U.S. Pat. 3,152,951** (Oct. 31, 1964), S. D. Pedman.
150. **U.S. Pat. 4,128,633** (Dec. 5, 1978), D. H. Lorenz and E. P. Williams (to GAF Corp.).
151. A. P. Black, *J. Amer. Water Works Assoc.* **60**(1), 69 (1968).
152. S. C. Werner, ed., *The Thyroid*, Harper & Row Publishers, New York, 1962.
153. S. L. Chang and J. C. Morris, *Ind. Eng. Chem.* **45**, 1109 (1953).
154. A. P. Black, J. B. Lackey, and E. S. Lackey, *Amer. J. Pub. Health* **49**, 1060 (1959).
155. E. L. Nilson and R. F. Unz, *Appl. Environ. Microbiol.* **34**, 815 (1977).

Armin Lauterbach
Gustavo Ober
SQM Iodine

IODINE VALUE.

See CARBOXYLIC ACIDS; FATS AND FATTY OILS.

IDOACETIC ACID.

See ACETIC ACID.

IDOFLUOROHYDROCARBONS.

See FLUORINE COMPOUNDS, ORGANIC.

ION EXCHANGE

Ion exchange is a process in which cations or anions in a liquid are exchanged with cations or anions on a solid sorbent. Cations are interchanged with other cations, anions are exchanged with other anions, and electroneutrality is maintained in both the liquid and solid phases. The process is reversible, which allows extended use of the sorbent resin before replacement is necessary.

Many naturally occurring inorganic and organic materials have ion-exchange properties. Synthetic organic ion-exchange resins became available in the late 1930s with the introduction of phenolic-type products. Styrenic resins appeared in the mid-1940s, and acrylic resins about 20 years later. The ion-exchange market of the early to middle 1990s is dominated by the styrenic resins, but acrylic resins are becoming increasingly important. Phenolic-based resins have almost disappeared. A few other resin types are available commercially but have not made a significant impact. Inorganic materials retain importance in a number of areas where synthetic organic ion-exchange resins are not normally used. Only the latter are discussed here. This article places emphasis on the styrenic and acrylic resins that are made as small beads. Other forms of synthetic ion-exchange materials such as membranes, papers, fibers (qv), foams (qv), and liquid extractants are not included (see EXTRACTION, LIQUID-LIQUID; MEMBRANE TECHNOLOGY; PAPER,).

The primary application for ion exchange is the softening and deionization of water (qv). The remaining applications include waste treatment (qv), catalysis (qv), purification of chemicals, plating, hydrometallurgy, food processing (qv), and pharmaceutical uses. Because ion-exchange resins are insoluble polymeric acids and bases, these resins are also useful in removing acids and bases from gaseous streams via the neutralization of functional groups.

Weak and strong acid-type resins are for removal of cations and are called cation exchangers. Weak and strong base resins remove anions and are called anion exchangers. In addition to these four resin types, there are specialty resins used in applications where higher specificity for certain ions under challenging conditions is a critical factor.

Continuous columnar operation of ion-exchange systems is preferred over batch operation. Each column must be taken off-stream periodically to remove the adsorbed ions and restore the resin to the ionic form required for the adsorption (qv) step. In this sense, a columnar ion-exchange operation is not continuous. Installations are usually designed with multiple units to assure a continuous flow of the process stream when one or more of the columns require regeneration. In those installations where the ion-exchange system is not required throughout the day, regeneration is scheduled during idle time and the system requires fewer ion-exchange units. Continuous operation has been approached in a number of designs by moving resin, or vessels containing resin, in a direction opposite to the flow of liquid. Some of these approaches have been abandoned; others are increasing in popularity.

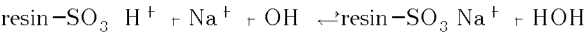
All aspects of ion exchange covered herein are presented in much greater detail in the numerous books devoted to the subject (1–7).

Types of Ion-Exchange Resins

Ion-exchange resins are categorized by the nature of functional groups attached to a polymeric matrix, by the chemistry of the particular polymer in the matrix, and by the porosity of the polymeric matrix. There are four primary types of functionality: strong acid, weak acid, strong base, and weak base. Another type consists of less common structures in specialty resins such as those which have chelating characteristics.

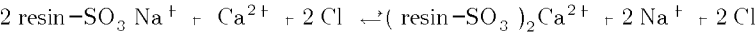
Cation-Exchange Resins.

Strong acid. Strong acid cation-exchange resins have sulfonic acid groups, -SO₃H⁺, attached to an insoluble polymeric matrix. When the functional groups are in the hydrogen form and the resin is in contact with a liquid containing other cations, hydrogen ions leave the solid phase and enter the liquid phase as they are replaced by cations from the liquid phase, for example,



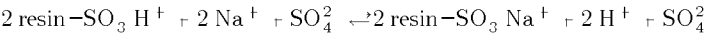
The liquid phase is free of Na⁺ and the functional groups of the resin are converted to a sodium salt. Multivalent cations are removed in a similar manner. Electric charge neutrality must be maintained in both the liquid and solid phases.

It is not always necessary for the resin to be in the hydrogen form for adsorption of cations, especially if a change in the pH of the liquid phase is to be avoided (see also HYDROGEN ION ACTIVITY). For example, softening of water, both in homes and at industrial sites, is practiced by using the resin in the ⁺ form.



Sodium ions are displaced from the resin by calcium ions, for which the resin has a greater selectivity.

In many industrial applications, strong acid cation-exchange resins are used in the hydrogen form to process liquids containing low concentrations of salts.



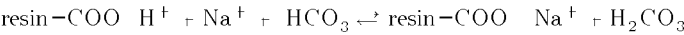
This is commonly referred to as a salt splitting reaction. The resin's selectivity for Na⁺ is greater than it is for H⁺. Anions are removed in a similar manner with an anion-exchange resin.

Ion-exchange reactions are reversible. A regeneration procedure restores the resin to the ionic form it was in prior to the adsorption step.

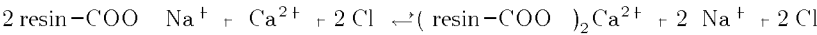
Reversibility of reactions allows resins to be used many times before replacement is considered. Strong acid cation exchangers are returned to the hydrogen, H⁺, form with dilute hydrochloric acid [7647-01-0] or sulfuric acid [8014-95-7]. Other mineral acids are used at times. However, the safety, cost, and methods of disposal must be thoroughly reviewed before using other acids. A 4% acid concentration is common. The use of higher or lower concentrations is dependent upon the process, the design of the system, and the potential for forming insoluble salts of the acid.

Weak Acid. Weak acid cation-exchange resins have carboxylic acid groups, -COOH attached to the polymeric matrix. Although not as versatile in process applications as the strong acid resins, these resins are included in numerous systems where higher operating capacities and greater ease in regeneration can be used advantageously.

Weak acid cation exchangers have essentially no ability to split neutral salts such as sodium chloride [7647-14-5]. On the other hand, an exchange is favorable when the electrolyte is a salt of a strong base and a weak acid.



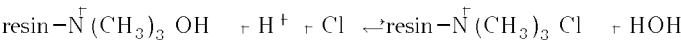
The sodium form of weakacid resins has exceptionally high selectivity for divalent cations in neutral, basic, and slightly acidic solutions.



The selectivityis so great that reversal of the reaction to restore the resin to the Na⁺ form is not practical using NaCl solutions at any concentration. Regeneration with dilute acid, followed by conversion of the resulting H⁺ form to the Na⁺ form with dilute sodium hydroxide [1310-73-2], is the preferred alternative.

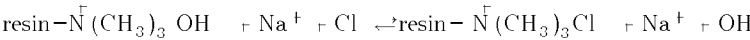
Anion-Exchange Resins.

Strong Base. Strong base anion-exchange resins have quaternary ammonium groups, -NR₃OH⁺, where R is usually CH₃, as the functional exchange sites (see QUATERNARY AMMONIUM COMPOUNDS). These resins are used most frequently in the hydroxide form for acidity reduction.

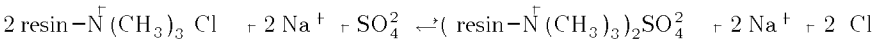


Hydroxide ions [14280-30-9] are released by the resin as anions are adsorbed from the liquid phase. The effect is elimination of acidity in the liquid and conversion of the resin to a salt form. Typically, the resin is restored to the OH⁻ form with a 4% solution of NaOH.

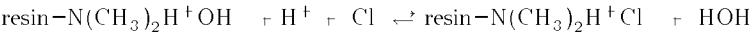
The hydroxide form is also used in salt splitting applications.



Salt forms of a strong base anion exchangerare used to remove other anions for which the resin has greater selectivity.

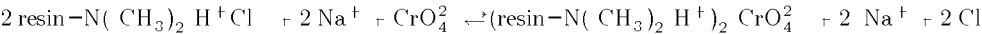


Weak Base. Weak base anion-exchange resins may have primary, secondary, or tertiary amines as the functional group. The tertiary amine -N(CH₃)₂ is most common. Weak base resins are frequently preferred over strong base resins for removal of strong acids in order to take advantage of the greater ease in regeneration.



Most weak base anion exchangers adsorbweak organic acids such as formic acid [64-18-6] and acetic acid [64-19-7], but do not remove weak organic acids such as carbonic acid [463-79-6] or silicic acid [7669-41-4].

Weak base resins when in the free base (hydroxyl) form are not capable of splitting neutral salts such as sodium chloride. Salt forms of weak base resins release anions to the liquid phase if other ions for which the resin has a greater selectivity are present.



This interchangeof ions is similar to that of the strong base resins.

Chromatographic Resins. The ion-exchange reactions illustrated are typical of those which occur in numerous industrial installations when the primary objective is the removal of ions, acids, or bases from a liquid stream. These same resins are useful for chromatographically separating ions having the same valence. Ion-exchange resins have a different selectivity for each ion. Sometimes these differences are too small for a separation to occur. In addition, other factors including flow rate, column design, and resin properties usually govern the success or failure of a chromatographic separation (see CHROMATOGRAPHY).

Although not of commercial interest, consider a separation of Ca²⁺ and Mg²⁺. A solution containing low concentrations of each cation is fed slowly and continuously to a column containing a strong acid-type cation-exchange resin in the Na⁺ form. At first the solution leaving the column contains only Na⁺ because it is displaced from the resin as Ca²⁺ and Mg²⁺ are adsorbed. As flow continues, Mg²⁺ appears in the effluent, but not Ca²⁺. Because the selectivity of the resin is greater for Ca²⁺ than for Mg²⁺, Ca²⁺ displaces the Mg²⁺ previously adsorbed by the functional groups. As flow continues and the column becomes loaded, Ca²⁺ appears in the effluent at increasing concentrations along with Mg²⁺.

Chromatographic separations are not limited to ionic constituents. For example, glucose [50-99-7], C₆H₁₂O₆, is separated from fructose [57-48-7], C₆H₁₂O₆, when using a strong acid cation exchanger in the Ca²⁺ form. The functional groups hinder the forward movement of each sugar at a slightly different rate as a solution containing both sugars flows slowly through a column containing the resin. Water [7732-18-5] is fed alternately with the sugar solution to aid in developing the separation. Glucose precedes the appearance of fructose in the effluent (8), indicating fructose is prevented in its forward motion by the resin to a greater degree than is glucose.

Resins having other types of functional groups are growing in importance. Resins that have metal chelating capabilities include those containing iminodiacetic acid [142-73-4] or aminophosphonic acid sites (9). Resins having thiol functionality are interesting for adsorption of metals which form sulfide precipitates. Resins containing an N-methylglucamine [6284-40-8] functionality are selective for boron [7440-42-8].

Manufacture of Resins

The production of ion-exchange resins is a multiple step process. It begins with the polymerization of monomers to form solid intermediate copolymers that are insoluble in both water and solvents. The copolymers are functionalized during additional steps in different reactors from those used for copolymer production. Conversion to another ionic form may be required after functionalization is completed. All resins are thoroughly rinsed with water to cleanse them of residual chemicals. Excess water is removed by vacuum filtration prior to packaging. Complete removal of all water by drying is unusual. Packaging in fiber drums is common. Alternative containers include metal drums, bulk boxes and bags, and smaller plastic, paper, and budap bags. Polyethylene [9002-88-4] liners are used as a barrier between containers and water-containing resin.

Manufacture of ion-exchange resins has traditionally been a batch process. Significant progress was made more recently in the development of a continuous process for the manufacture of copolymer beads. However, as of this writing (ca 1994) is it not used by all manufacturers. Moreover, those companies having continuous processing capabilities do not use it for all ion-exchange products.

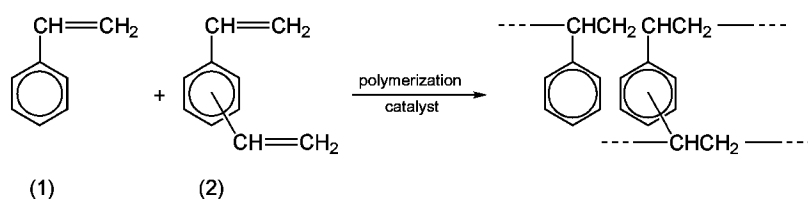
Copolymerization. The chemistry of the resin matrix, the type and degree of porosity, the particle size, and the particle size distribution are established in the copolymerization step. Formulations and operating procedures must be strictly followed. Reaction vessels must be well designed. Mistakes made during copolymerization are rarely corrected during functionalization.

The procedure of forming copolymers dates back to the early 1940s when only phenolic resins were available. Copolymers were produced by bulk polymerization of phenol [108-95-2] and formaldehyde [50-00-0]. Because the resulting solid product had the shape of the vessel in which polymerization

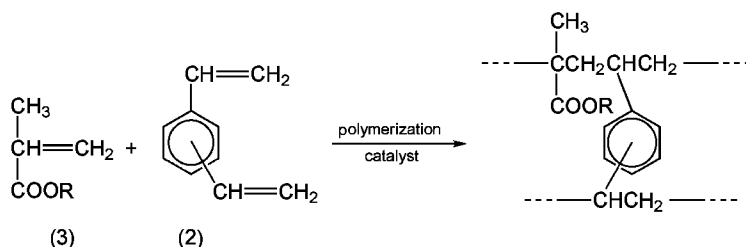
took place, it had to be reduced to smaller particles by crushing and grinding before being functionalized. A new chemistry, and a new process, appeared in the mid-1940s when it was learned that styrene [100-45-2] could be copolymerized with divinylbenzene [1321-74-0] (DVB) by a suspension polymerization process, and that the small, hard copolymer beads could be functionalized (10,11). These styrenic resins eventually replaced the phenolic resins in practically all applications. Since then, another chemistry has evolved to complement the styrenic resins in process applications. These are the acrylic resins which are functionalized copolymers of an acrylic monomer and DVB produced by the suspension polymerization technique.

Design of copolymer reactors became more complex with the introduction of suspension polymerization. The size and uniformity of the ion-exchange resin to be produced from a copolymer was now dependent on the size and uniformity of a liquid monomer mix dispersed as small droplets in an aqueous medium. The shape, size, and speed of mixers, as well as baffling to control fluid flow within the reactors, became critical factors. These reactors must be closed to the atmosphere, and they must be fitted with appropriate piping and valves to allow rapid transfer of organic and aqueous phases from other tanks to the reactor, and to allow rapid removal of a copolymer slurry when the reaction is over. The reaction kettles are jacketed which provides a means for raising and lowering temperatures during polymerization. Careful selection of materials of construction assures long life, minimizes downtime for maintenance, and guards against contamination of the liquid and aqueous phases with elements which may interfere with the preferred rate of polymerization.

The organic and aqueous phases are prepared in separate tanks before transferring to the reaction kettle. In the manufacture of a styrenic copolymer, predetermined amounts of styrene (1) and divinylbenzene (2) are mixed together in the organic phase tank. Styrene is the principal constituent, and is usually about 90–95 wt % of the formulation. The other 5–10% is DVB. It is required to link chains of linear polystyrene together as polymerization proceeds. DVB is referred to as a cross-linker. Without it, functionalized polystyrene would be much too soluble to perform as an ion-exchange resin. Ethylene-methacrylate [97-90-5], and to a lesser degree trivinylbenzene [1322-23-2], are occasionally used as substitutes for DVB.



Formulations for acrylic copolymers involve monomers such as acrylic acid [79-10-7], methacrylic acid [79-14-4], or esters of these acids. Formation of a copolymer from a methylmethacrylate ester (3), where DVB serves as the cross-linker, gives the structures:



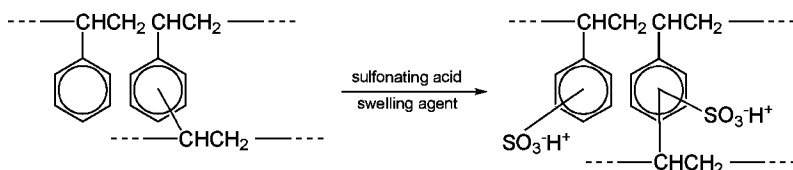
The fraction of DVB used in the monomer mix is governed by the required performance characteristics of the ion-exchange resin. The higher the percentage of DVB, the greater the number of sites at which polymeric chains of styrene, or acrylic monomer, are linked. High DVB levels increase the tightness of the polymer matrix and lessen the ability of ions or other solutes to migrate through the resin phase. Thus the porosity decreases as the cross-linking increases. When the organic phase contains only the monomers that participate in polymerization, the copolymers and resins produced from them are described as having a microporous or gel-type porosity. It is a porosity which cannot be measured by conventional methods. A different type of porosity, macroporous (also called macroreticular), is formed by incorporating a solvent along with the monomers in the organic phase. That solvent should have low aqueous phase solubility and must not participate in the polymerization reaction. When the organic phase is dispersed into the aqueous phase, the solvent is distributed throughout each droplet. The solvent, which remains with the hard copolymer beads during polymerization, is not bound to the resin matrix because it did not react with either monomer. Displacement of the solvent from the copolymer, or removal by vaporization, yields a copolymer with a measurable pore volume and pore size distribution. Both pore volume and distribution are dependent on the solvent and the amount incorporated in the original monomer mix.

The aqueous phase into which the monomer mix is dispersed is also prepared in a separate tank before transferring to the copolymerization kettle. It contains a catalyst, such as benzoyl peroxide [94-36-0], to initiate and sustain the polymerization reaction, and chemicals that aid in stabilizing the emulsion after the desired degree of dispersion is achieved. Careful adherence to predetermined reaction time and temperature profiles for each copolymer formulation is necessary to assure good physical durability of the final ion-exchange product.

Continuous processes for copolymer production were developed initially for the microporous resins. The system generally involves injecting the monomer mix into the aqueous phase through orifice plates. Droplet size is controlled by the diameter of the holes in the plate and the rate at which the monomer is injected into the aqueous phase. The continuous process produces copolymer beads which have greater uniformity in size than those produced in batches.

Functionalization. Copolymers do not have the ability to exchange ions. Such properties are imparted by chemically bonding acidic or basic functional groups to the aromatic rings of styrenic copolymers, or by modifying the carboxyl groups of the acrylic copolymers. There does not appear to be a continuous functionalization process on a commercial scale.

Strong Acid Cation Exchangers. All strong acid-type resins are made from styrene–DVB copolymers, with the exception of a minor quantity of phenolic resin. Batch sulfonation using commercial strength sulfuric acid [8014-95-1] is common.



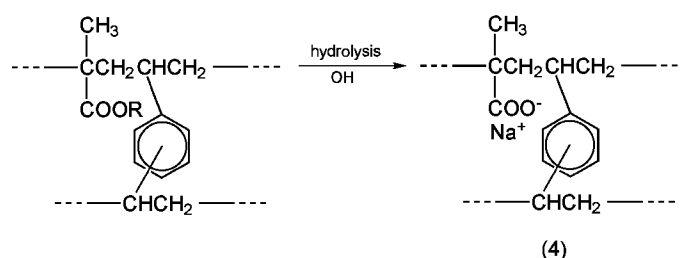
As in copolymerization, time–temperature profiles are followed closely in order to attach $\text{-SO}_3\text{H}$ groups to aromatic rings throughout the resin particle. Aromatic rings at the core of the beads are not as accessible as those closer to the surface. Sulfonation is usually carried out in a solvent such as ethylene dichloride [107-06-2] or propylene dichloride [78-87-5]. It can be accomplished without the solvent. Appearance and physical stability of sulfonated products produced in the presence of a solvent are generally superior to resins sulfonated in the absence of a solvent. On the other hand, more environmental problems are associated with sulfonation in the presence of solvents. Whichever process is used, very little acid is consumed relative to the total amount in the reactor. Separation of a highly acidic liquid from the resin after functionalization must be approached with care. This liquid can be displaced with an acid of lower concentration. Otherwise, a dramatic change in the acid concentration results in rapid swelling and subsequent cracking of the resin particles. Recycling or reuse of acids is limited by the lowering of the acid concentration and by the increase in concentration of organic

impurities.

Commercial demand for strong acid resins is greatest for those having microporous properties and a copolymer DVB content of 8%. Resins having greater cross-linking are generally preferred in processes where significant oxidative attack is expected because these are more resistant to deterioration. Resins functionalized from copolymers with less than 8% DVB are used in a variety of nonwater treatment applications where a more open polymeric matrix provides a better pathway for large ions and compounds migrating to functional groups within the resin particles. Macroporous cation exchangers usually have 12 to 25% DVB. They are particularly useful in applications where oxidative attack is severe. They are also beneficial in nonaqueous systems. A substantially larger number of functional groups, or exchange sites, are available along the internal surfaces present in macroporous-type resins, compared to just those sites on the outer surface of an unswollen microporous resin.

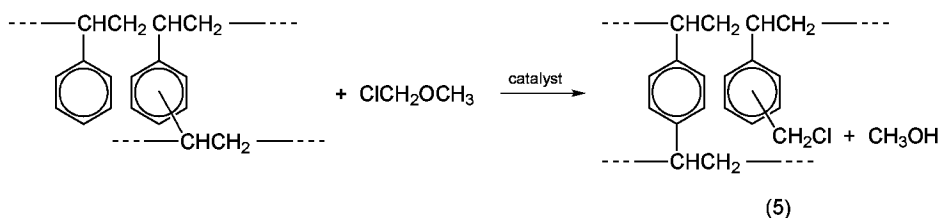
Corrosion of reactors used for functionalization and in pipes and valves along transfer lines for sulfuric acid is a problem that results in maintenance shutdowns. Sufficient agitation is needed to keep the resin beads fluidized during sulfonation. As for copolymer kettles, transfer lines should be sufficiently large to allow reasonably rapid transfer of liquids and resin slurries.

Weak Acid Cation Exchangers. The synthesis of weak acid cation exchangers is a one-step process when acrylic acid or methacrylic acid is copolymerized with DVB. If an acrylic ester is used as the monomer instead of an acrylic acid, the ester groups must be hydrolyzed after polymerization using either an acid or base (NaOH) to give the carboxylic acid functionality, or the sodium salt (4) of it.



The market for weak acid cation exchangers is much less than that of strong acid resins. As a consequence, fewer resin variations are available commercially. Very little published information is available concerning the degree of cross-linking. Most weak acid resins have a macroporous structure, although the pore volume is significantly less than that of the strong acid resins.

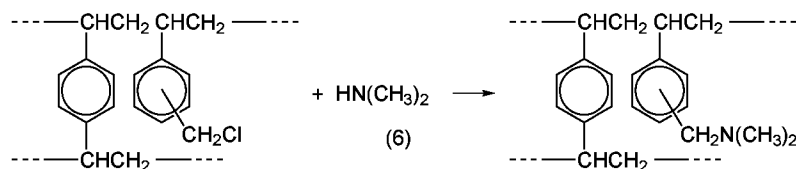
Weak Base Anion Exchangers. Both styrenic and acrylic copolymers can be converted to weak base anion-exchange resins, but different synthetic routes are necessary. Styrene-DVB copolymers are chloromethylated and aminated in a two-step process. Chloromethyl groups are attached to the aromatic rings (5) by reaction of chloromethyl methyl ether [107-30-2], $\text{CH}_3\text{OCH}_2\text{Cl}$, with the copolymer in the presence of a Friedel-Crafts catalyst such as aluminum chloride [7446-70-0], AlCl_3 , iron(III) chloride [7705-08-0], FeCl_3 , or zinc chloride [7646-85-7], ZnCl_2 .



The presence of bis(chloromethyl) ether [542-88-1], $\text{ClCH}_2\text{OCH}_2\text{Cl}$, in chloromethyl methyl ether generally results in additional secondary cross-linking when the chloromethyl groups at each end of the molecule combine with aromatic rings on two separate polystyrene chains. The amount of secondary cross-linking is dependent on the catalyst and the reaction parameters. As a consequence, copolymers synthesized specifically for the manufacture of anion exchangers tend to have a lower DVB content than copolymers used in the production of strong acid cation exchangers since the cumulative cross-linking has an effect on the performance properties of the ion exchanger.

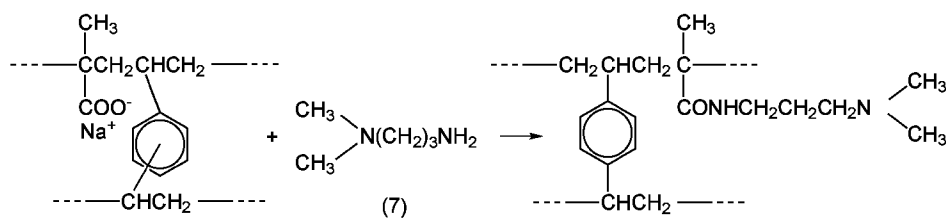
The chloromethylation reaction and the steps associated with its preparation and transfer to the reaction vessel are the most hazardous procedures in the manufacture of any ion-exchange resin. Bis(chloromethyl) ether and chloromethyl methyl ether are recognized carcinogens. Chloromethyl methyl ether is manufactured on site from methanol [67-56-1], or methylal [109-87-5], and formaldehyde. Some ion-exchange resin plants have a separate facility. Others produce it in the same kettle used for the chloromethylation reaction. Leaks in pipe joints, valves, and reaction vessels cannot be tolerated. Most if not all resin manufacturers have converted older processing facilities to fully automated systems. Buildings in which the reactions occur are placed under negative pressure. Personnel are not permitted in the production building unless protective masks and clothing are worn. Bis(chloromethyl) ether and chloromethyl methyl ether remaining after completion of the chloromethylation reaction are recycled or completely destroyed before the resin is aminated.

Functionalization is completed by aminating the chloromethylated copolymer with either primary or secondary amines. Dimethylamine [124-40-3] (6) is generally preferred, especially in the synthesis of the macroporous anion exchangers.



Polyamines, such as diethylenetriamine [111-40-0] are used at times in the synthesis of microporous weak base resins to achieve significantly higher capacity. However, these resins generally have lower physical and chemical stability than resins prepared from primary or secondary amines.

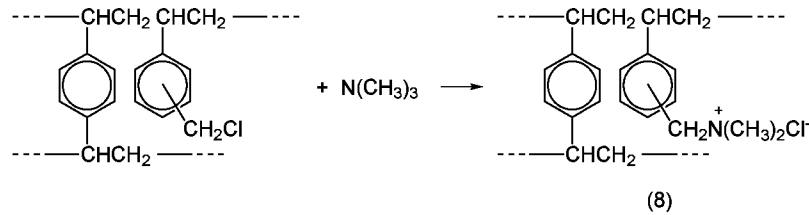
The acrylic weak base resins are synthesized from copolymers similar to those used for the manufacture of weak acid cation-exchange resins. For example, under appropriate temperature and pressure conditions, a weak acid resin reacts with a polyfunctional amine, such as dimethylaminopropylamine [109-55-7] (7) to give a weak base resin with a tertiary amine functionality.



Both the styrenic and the acrylic weak base resins are used in industrial applications for the same purposes, primarily the removal of acidic components in

liquid streams. The styrenic weak base resins are more stable at temperatures over 37°C than the acrylic resins. On the other hand, the acrylic resins are more hydrophilic and tend to adsorb and release large molecular weight constituents more effectively than the styrenics. Levels of cross-linking are not usually disclosed by resin manufacturers. The variety of resins available is not as great as for the strong acid cation exchangers.

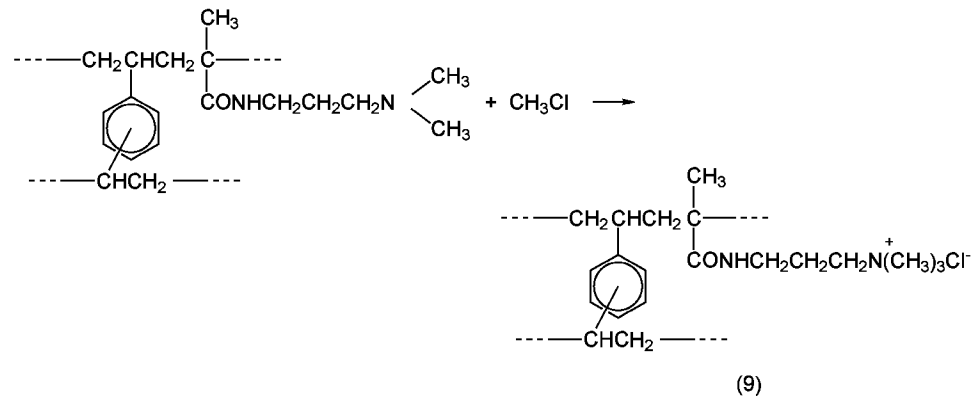
Strong Base Anion Exchangers. As in the synthesis of weak base anion exchangers, strong base resins are manufactured from styrenic as well as acrylic copolymers. Those based on copolymers of styrene and divinylbenzene are chloromethylated and then aminated. These reactions are the same as for the styrenic weak base resins. The essential difference is the amine used for amination. Trimethylamine [75-50-3], N(CH₃)₃, and N,N-dimethylethanolamine [108-01-0], (CH₃)₂NCH₂CH₂OH, are most commonly used. Both form quaternary ammonium functional groups similar to (8).



Styrenic resins with this structure are commonly called Type I strong base anion exchangers (8) in the water treatment industry. Those aminated with dimethylethanolamine are referred to as Type II anion exchangers. These groups are slightly less basic than those of the Type I resins. As a consequence, anions are not held as tightly and they are replaced more efficiently by hydroxide ions when regenerated with NaOH solutions. Type II resins have lower chemical and thermal stability than Type I resins. Higher molecular weight amines such as triethylamine [121-94-8] and tributylamine [102-82-9] are used occasionally to make strong base resins which have enhanced selectivity for ions such as nitrate [14797-55-8], NO₃⁻, over other ions normally present in water supplies and wastewaters. However, the mobility of ions through the resin matrix and the exchange capacity decrease as the size of the amine used for amination increases.

Low molecular weight amines are volatile and extremely odoriferous. They are removed from the atmosphere of the manufacturing facilities by scrubbing to comply with environmental regulations. An amine odor is common with packaged anion-exchange resins even though the resins are thoroughly washed before pack-out. Odor is most common for strong base resins aminated with trimethylamine, and when in the hydroxide form.

Acrylic strong base anion exchangers (9) are synthesized from acrylic weak base resins. The tertiary amine groups are converted to a quaternary ammonium functionality by reaction of chloromethane [74-87-3], CH₃Cl, and the weak base resin.



Referring to this quaternary structure in acrylic resins as Type I is controversial. Many of the properties are similar to those of the Type II styrenic resins. Precautions regarding chemical and thermal stability are generally the same for acrylic strong base resins and the Type II styrenics. In addition, the ability to remove adsorbed ions during regeneration is about the same for the acrylic resins and Type II resins. This supports a suggestion that the basic strength of the $-\text{N}(\text{CH}_3)_3$ group on the acrylic is not as strong as the same group on the styrenic Type I resin, but more like the $-\text{N}(\text{CH}_3)_2\text{CH}_2\text{CH}_2\text{OH}$ on the Type II resins.

Acrylic anion exchangers are frequently preferred over the styrenic resins for removal of high molecular weight organic acids present in surface water supplies and in other process streams. Compounds such as these are not completely removed during the regeneration step. The accumulation of these compounds with increased cycles of use is referred to as organic fouling. The more hydrophilic structure of the acrylic resins provides a less complex pathway for the organic compounds to migrate out of the resin during regeneration, and organic fouling is minimized.

A simplified schematic layout of an ion-exchange production facility is presented in Figure 1. Layouts vary from one company to another and are significantly more complex when recycle of streams and environmental controls are incorporated in the schematics.

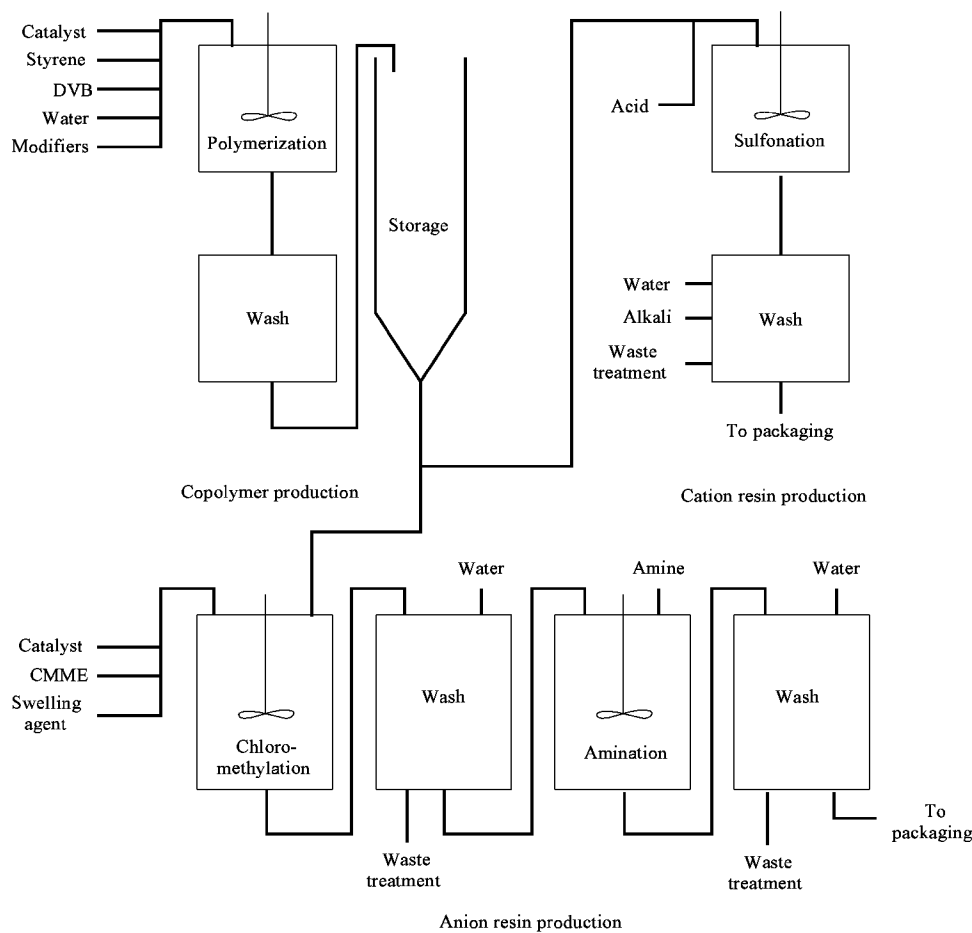


Fig. 1. Ion-exchange production schematic. CMME is chloromethyl methyl ether.

Separate kettles and backwash towers are frequently used to convert ion-exchange resins from one ionic form to another prior to packaging, and to cleanse the resin of chemicals used in the functionalization reactions. Excess water is removed from the resin prior to packaging by a vacuum drain. Both straight line filters and towers or columns are used for this purpose.

Physical and Chemical Properties

Ion-exchange resins are used repeatedly in a cyclic manner over many years, and deterioration of both physical and chemical properties can be anticipated. Comparison of the properties of used resin with those of new resin is helpful to learning more about the nature and cause of deterioration (12). Corrective action frequently extends the life of the resin. Comparison of properties must always be made with the resin in the same ionic form.

Particle Shape and Size. With few exceptions, resins are supplied as small, round beads having a diameter between 0.3 and 1.2 mm. Some resins are reduced to a smaller size by grinding to satisfy specific requirements in applications for electric power generation (qv) and pharmaceuticals (qv).

Resin size is dependent on two factors. One of these is the method by which the copolymer beads are formed, and the other is related to the exchange kinetics within the beads. Copolymer beads are made by batch-wise suspension of a liquid monomer mix in an aqueous medium or by a continuous jetting technique. In either case, formation of beads larger than 1 mm in diameter is difficult to achieve without agglomeration. Large beads are preferred for pressure drop advantages in columnar operations. However, large beads are subject to a greater rate of breakage than those having a smaller size.

Migration of ions to functional groups within the resin particles is slow compared to the aqueous or organic liquid phase surrounding the particles. The most efficient ion-exchange processes occur when most of the functional groups can be accessed within a short contact time between liquid and resin. The larger the particle size, the greater is the time required to utilize groups deeper in the particles. Thus, the smaller the bead, the better. However, the smaller the size, the greater the pressure drop. A particle size range spanning 0.3–1.2 mm in diameter has been a compromise between acceptable kinetics and pressure drop.

A Gaussian distribution of particle size is the result of copolymer manufactured by suspension polymerization. A jetting process produces beads with more uniform particle size. The uniformity coefficient is a numerical method of indicating closeness of all beads to the same size.

Wet screening using a set of U.S. Standard sieves (297, 420, 595, and 841 μm openings; 50, 40, 30, and 20 mesh, respectively) is the standard procedure for determining particle size distribution. The volume percent retained on each is recorded. The cumulative percent retained is plotted as the abscissa and the screen opening in mm as the ordinate on log–log probability graph paper. The best straight line is determined. The uniformity coefficient is defined as the ratio of the screen opening that retains 90% of the particles to the opening that retains 40%. The effective size is defined as that screen opening which retains 90% of the particles, as obtained from the same graph. If both the uniformity coefficient and effective size are available for a batch of resin, the approximate distribution by sieve opening can be calculated.

Wet screening is being replaced in many laboratories by instruments that use a photosensor to record the diameter of each particle in a water suspension as that suspension flows past the sensor.

Density and Specific Gravity. Density generally pertains to the bulk, or pack-out, weight of wet resin per unit volume. The density is characteristic of the resin and is dependent on the copolymer structure, the degree of cross-linking, the nature of the functional groups, and the ionic form of those groups. A change in density after extended use is a signal that chemical degradation has occurred. The density of most cation exchangers is in the 800–900 g/L range, whereas most anion exchangers are in the 640–740 g/L range.

The specific gravity generally refers to the value determined for wet resin when using a pycnometer. Values range from about 1.04 to about 1.25. Cation exchangers have a greater specific gravity than anion exchangers.

Porosity. The structure of ion-exchange resins is either microporous or macroporous. Microporous resins are more commonly referred to as gel or gelular-type resins. Porosity of this type resin cannot be measured by standard techniques. Gel resins are porous when the particles are swollen with water or another solvent. There is no porosity when the resin is dry. Microporosity is inversely proportional to the degree of cross-linking. Large ions migrate through a low cross-linked resin faster than through the less porous, higher crosslinked resins.

Macroporous resins are also called macroreticular. Macroporous resins have a measurable porosity. It does not disappear when the resin is dry. Porosity is more dependent on the solvent used when manufacturing the copolymer than on the degree of cross-linking.

Capacity. Capacity is a measure of the quantity of ions, acid, or base removed (adsorbed) by an ion-exchange material. The quantity removed is directly correlated with the number of functional groups. Capacity is reported in several different ways, but requires further definition because the word by itself does not cover all situations. Total capacity is a measure of all the functional groups on a resin and is recorded on a weight as well as a volume basis. A manufacturing objective is to place at least one functional group on each aromatic ring in styrenic-type resins. As a consequence, the degree of cross-linking has little effect on the total dry weight capacity, as shown in Figure 2, for strong acid cation exchangers in the hydrogen form. Dry resins swell when wet with water. The amount of swelling decreases as the degree of cross-linking increases. For this reason, the effect of cross-linking on total volumetric capacity is more demonstrative, as is also shown in Figure 2 for the same resins.

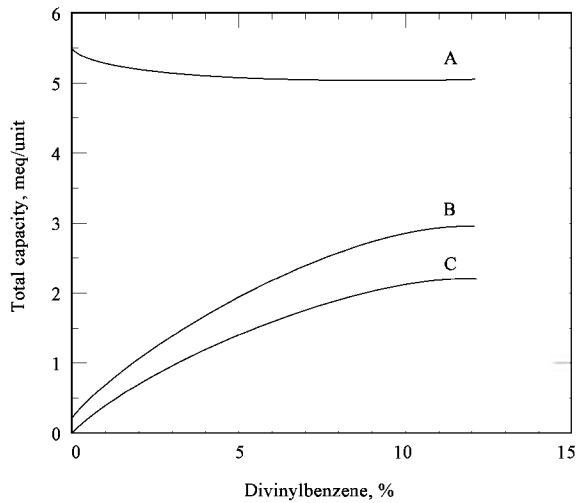


Fig. 2. Total capacity vs cross-linkage for polystyrene sulfonic acid resin in the H⁺ form where A and B correspond respectively to dry and wet weight capacity in meq g of resin, and C represents wet volume capacity in meq mL of wet settled resin.

Operating capacity, also called the working capacity or column capacity, is a measure of the quantity of ions, acids, or bases adsorbed, or exchanged, under the conditions existing during batch or columnar operation. Because the adsorption process is terminated in most commercial applications before all functional groups have been utilized, the operating capacity is less than the total capacity. Operating capacities vary from one installation to another, even though the same resin might be used at each location. Such variations are the result of differences in composition of the stream to be treated, flow rate, effluent quality that triggers shutdown for regeneration, and regeneration conditions. Examples of the dependence of operating capacity on the degree of regeneration and resin type are illustrated in Figure 3 for typical weak base and strong base anion exchangers.

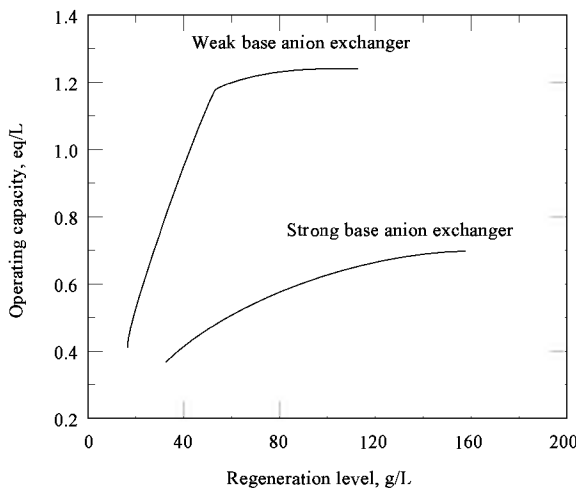
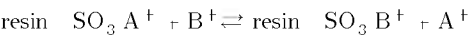


Fig. 3. Regeneration of anion-exchange resins using NaOH (14).

The steeper slope for the weak base resin is an indication that these resins are regenerated with greater efficiency, or ease, than strong base resins. Strong base anion exchangers do not release anions as readily as weak base anion exchangers. For each incremental increase in the amount of chemical used for regeneration there is a greater incremental increase in the operating capacity of the weak base resin than the strong base resin. If two resins have the same total capacity but differ in basic strength and are regenerated under identical conditions, the resin having the lower basicity is restored more completely to a regenerated form and has the higher operating capacity. A similar analogy can be made for weak and strong acid cation exchangers.

Selectivity. A significant exchange of ions does not occur unless the functional group of the resin has a greater selectivity for ions in solution than for ions occupying the functional group, or unless there is a mass action effect, as in regeneration. Selectivity coefficients have been reported in numerous publications for both cations and anions. These coefficients are determined at very low concentrations using the two specific ions in question. In the following reaction, for example, the cation-exchange resin removes cationsB⁺ from solution in exchange forA⁺ on the resin because the selectivity of the resin forB⁺ is greater than forA⁺.



At equilibrium, the selectivity coefficient K_A^B for B over A is determined from the following equation:

$$K_A^B = \frac{\overline{m}_B \cdot m_A}{\overline{m}_A \cdot m_B}$$

where m and $\overline{m}$ are the ionic concentrations of the ions in the solution and resin phase, respectively. Selectivity for ions having the same charge usually increases as the atomic weight increases (Li<Na<K<Rb<Cs and Mg<Ca<Sr<Ba). Selectivities for divalent ions are greater than for monovalents. Higher selectivity for trivalent and tetravalent ions does not necessarily follow the expected progression. The following examples (Table 1) have been chosen to illustrate the effect of type of strong base functionality and mole fraction of chloride [16887-00-6] in a two-component anion system on selectivity constant (13).

Table 1. Selectivity for Anions on Anion Exchangers ^a

Ion	CAS Registry Number	Dowex 1 ^b		Dowex 2 ^c		Dowex 2 ^c	
		X _{Cl}	K _{Cl}	X _{Cl}	K _{Cl}	X _{Cl}	K _{Cl}
iodide	[20461-54-5]	0.27	8.7	0.27	7.3	0.07	13.2
nitrate	[14797-55-8]	0.38	3.8	0.36	3.3	0.34	3.3
nitrite	[14797-65-0]	0.51	1.2	0.52	1.3		
hydroxide	[14280-30-9]	0.77	0.09	0.56	0.65		
bicarbonate	[71-52-3]	0.65	0.32	0.63	0.53		
formate	[71-47-6]	0.70	0.22	0.68	0.22		
fluoride	[16984-48-8]	0.77	0.10	0.76	0.10		

^a Selectivity is vs that for Cl⁻. Mol fraction of Cl⁻, X_{Cl}, is given.

^b Type I functionality.

^c Type II functionality.

The need to know selectivity coefficients precisely is rarely necessary in industrial applications. However, knowledge of relative differences is important when deciding if the reaction is favorable or not. Most ion-exchange applications are for the removal of more than one ionic species. In water softening, it is important to know that a cation exchanger has a greater selectivity for divalent cations than for monovalent cations. It is not critical to know the resin prefers Ca²⁺ over Mg²⁺. In deionization, it is important to know that all cations normally present in a water supply are preferred by a strong acid cation exchanger over a hydrogen ion [12408-02-5], H⁺, and that all anions are preferred over OH⁻ by a strong base anion exchanger. Differences in selectivity coefficients take on greater importance in chromatographic separations where flow rates are much slower to allow more time for ionic species to separate as the liquid flows through the resin bed.

Selectivity differences increase as the degree of cross-linking of a resin increases, but these differences are relatively minor. Structural composition of the functional groups has a much greater effect on the magnitude of selectivity differences, as is illustrated for softening of water supplies. The sodium form of either a strong or a weak acid cation exchanger removes Ca²⁺ and Mg²⁺ from water effectively. However, the selectivity of the weak acid resin is so much greater that very concentrated solutions of NaCl do not reverse the reaction during the regeneration step, as it does for the strong acid cation exchanger. Even greater selectivity for divalent cations is observed for resins having aminophosphonic acid or iminodiacetic acid functionality.

Kinetics. The degree to which an ion-exchange reaction is completed depends on a number of factors which include contact time, ionic concentration, degree of cross-linking, and temperature. Contact time and concentration are interrelated and most important for a successful ion-exchange process. The impact of temperature increases as the viscosity increases. In columnar operations, short contact times associated with high flow rates are acceptable when the concentration of ionic constituents in the influent stream is very low. At higher electrolyte concentrations, the contact time must be increased. If contact time is too short, functional groups on or near the surface are converted to another ionic form sooner than groups deeper in the resin particles. The net result is the appearance of ions that were to be adsorbed in the effluent sooner than would have occurred with a longer contact time, and lower utilization of all the resin's capacity. Flow rates of 8 to 40 bed volumes (BV) h (1–5 gal min⁻¹ ft⁻³) are common in conventional water treatment systems where ionic concentrations might be as high as several hundred mg L⁻¹. Slower flow rates should be considered for ionic concentrations exceeding 500 mg L⁻¹. For those systems which have ionic concentrations in the μg L⁻¹ range, as in condensate polishing and similar recycle applications, flow rates are usually significantly greater than 40 BV h.

A hypothetical illustration of the ion concentration–flow rate effect is presented in Figure 4. Consider two identical columns operating under the same conditions, except for the ionic concentration of the influent solution. One stream, curve C, has a concentration x; and the other 2x. Once ions appear in the effluent, they increase in concentration as more solution passes through. These curves, showing effluent concentration vs the volume of solution treated, are called leakage curves. For purposes of this example, the column is shut down for regeneration when the effluent concentration is 5% of the influent concentration. If the stream having twice the concentration had no effect on the exchange rate, then the leakage curve would be represented by curve B. Only half the volume could be handled by the resin because the ionic concentration is twice as much. The operating capacities for columns having leakage curves represented by curves C and B would be the same. If the flow rate is too great to allow the migration of ions to exchange sites deeper in the resin particles, then a leakage curve such as curve A can be expected. The operating capacity is less in this example because the adsorption cycle is terminated when ions in the effluent reach 5% of the influent concentration. The lower capacity is represented by the difference in the volume processed to the 5% cut-off point by curves B and A.

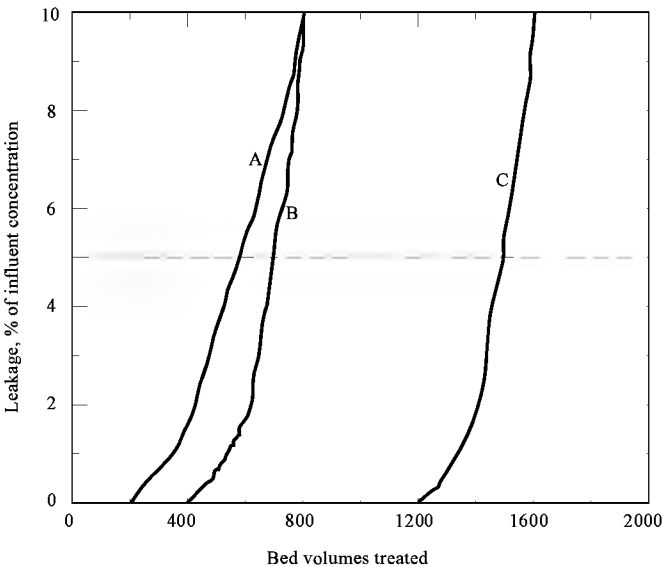


Fig. 4. Hypothetical leakage curves to show effect of ionic concentration and flow rate, where the dashed line corresponds to the leakage shutdown level for regeneration. See text.

Some industries practice ion exchange in nonaqueous systems. These solvents may cause resin particles to shrink or swell. Shrinkage has a negative effect on the kinetics, whereas swelling opens up the structure and improves migration of those constituents to be adsorbed. Microporous resins usually do not work well in nonaqueous systems because of the disappearance of porosity. Macroporous resins, however, are more satisfactory in these systems since porosity is retained even if the resins are dried completely. More functional groups on outer and inner surfaces are available for exchange as a result of the

combined fixed porosity and relatively high surface area. Nevertheless, these systems are operated at slower rates to compensate for slower migration of ions, acids, and bases through organic solvent or process stream surrounding the resin particles.

Several aqueous systems should be considered in a similar manner. For example, in the selective removal of divalent cations from a saturated salt solution, the hydrated resin gives up a portion of its normal water content as it contacts the salt stream. In so doing, the particles shrink, and the inner pathways for ion migration become smaller.

Moisture and Water Content. Resins are thoroughly washed with water upon completion of manufacture and conversion (if necessary) to another ionic form. Excess water is removed by vacuum draining or filtration. Nevertheless, a significant quantity of water associated with the functional groups and adhering to the outer surface of the resin particles remains with the resin as it is discharged into shipping containers. No effort is made to dry the resin, except in a few application areas, since the resins are used in aqueous processes in most installations.

Each resin has a characteristic water content dependent on the resin matrix, the structure of the functional groups, and the ionic form of those groups. Resins are packaged by weight and sold by volume. The dewatering operation prior to packaging is a critical step since removal of too much water is costly to the manufacturer, and removal of too little is costly to the buyer. Analyzing for water content is important to both the seller and user. The quantity of water contained by the resin is recorded on a percentage basis and determined by two methodologies. In each procedure, a small (ca 15 g) sample is removed from a larger composite sample collected during pack-out. In one procedure, the sample is accurately weighed before and after placing in a 105°C oven for at least 8 h. This procedure yields the moisture content typical of resin contained in the shipping containers. In the other procedure, a similar sample is soaked in water, then filtered under vacuum in a Buchner funnel prior to weighing before and after oven drying. The moisture content reported is a pseudoequilibrium value typical of the specific resin and its ionic form. If the value reported is either greater or lower than the expected range, all manufacturing steps need to be examined for deviation from the standard manufacturing procedure. Values determined by both methods rarely differ by more than 1%. A significant difference is an indication of a procedural change or malfunction of equipment in the plant dewatering step prior to packaging. Water content of strong base resins in the hydroxide form cannot be determined in the same manner since thermal degradation with additional weight loss occurs at 105°C.

Swelling and Shrinking. Ion-exchange resins shrink or swell reversibly as they are converted from one ionic form to another. The degree of change is dependent on the resin matrix, the functional group, and the ions adsorbed by the functional groups. For similar matrixes, the magnitude of volume changes decreases as the level of cross-linking increases. Resins in contact with solutions having a high electrolyte concentration shrink as water is drawn from the resin. These reswell when in contact with more dilute solutions. Wetting the resins with nonaqueous solvents causes shrinking or swelling, depending on the solvent. Swelling is greatest for weak acid cation exchangers (up to 100%) . Weak base anion exchangers may swell as much as 50% . Strong base resins swell about 15–25% . Strong acid-type resins swell the least and are usually in the 5– 10% range.

The degree of swelling and shrinking is important for design of ion-exchange columns, especially for the location of the distributors used to disperse incoming fluids, and collect outgoing ones, evenly over the cross-sectional area of the resin bed. Once placed, these distributors are not adjustable. The upper distributor should be above (the lower one below) the resin bed, even in the bed's swollen form.

Hydraulic Properties. Both the resistance to liquid flow through a resin bed and the degree to which a resin bed expands during a backwashing step are important design factors for ion-exchange systems. These characteristics are also critical to those using the resins because movements of resins not only signal the existence of a problem but give indications as to the nature of the problem. Pressure drop and hydraulic information for new resins are available from the resin manufacturer and the supplier of equipment.

Factors which have the greatest impact on pressure drop are the depth of the bed, flow rate, viscosity, temperature, and particle size. Pressure drop is also dependent on the hardness of the resin, or its ability to resist deformation when under compressive forces as in columnar systems with the process stream flowing down. Styrenic resins are more resistant to deformation than acrylics, and macroporous resins are more resistant than microporous. Minor increases in pressure drop as a result of bed packing are to be expected in lengthy downflow runs. Resins which have been used for a large number of cycles may yield a significantly higher pressure drop compared to those when the resin was new. These increases suggest problems with dirt, biological growth, or an increase in resin fragments within the column. Another cause may be oxidative attack of the resin which, in effect, makes it behave as a more compressible lower cross-linked resin.

Examples of pressure drop variation for new resin as a function of flow rate and water temperature are shown in Figure 5 for a standard styrenic strong acid cation exchanger. The lower pressure drop at the higher temperature is a reflection of water viscosity.

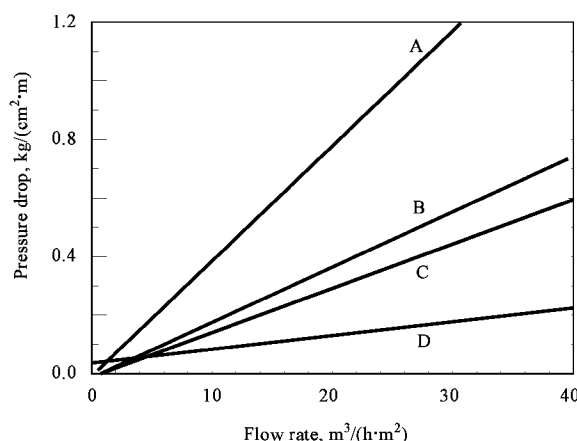


Fig. 5. Pressure drop as affected by resin type, flow rate, and temperature, where A, B, and C, correspond respectively to acrylic strong base anion exchanger (Amberlite IRA-458), styrenic strong base anion exchanger (Amberlite IRA-402), and styrenic strong acid cation exchanger (Amberlite IR-120), all at 4°C. D represents styrenic strong acid cation resin (Amberlite IR-120) at 50°C (14). To convert kg (cm²·m) to lb (in.²ft), multiply by 4.33; to convert m³ (h·m²) to gal (minft²), multiply by 0.409.

Backwashing is the upward flow of water through a bed of resin at a flow rate sufficient to fluidize the resin, but not so great that resin is carried out of the column with the exiting water. Resins are backwashed to remove dirt and resin fragments, to classify resin particles by size, and to relieve any packing that may have occurred with previous use. Frequency may be as great as once each cycle. Backwash times are 15 to 30 minutes, unless conditions require more time. Resin columns are designed with adequate space above the resin bed to allow 50–100% expansion during backwash. Each resin has a characteristic bed expansion profile which is dependent on the resin's specific gravity and particle size distribution. Severe accidental losses of resin occur during the colder months when the temperature of water used for backwashing is ignored. Lower temperature increases the water viscosity which increases the bouyancy effect on the resin particles. If the flow rate of the backwash water is not reduced, the bed expansion may be so great that resin particles leave the column with the exiting water. In warmer months, backwash flow rate should be increased. The dependence of expansion on flow rate and temperature is shown in Figure 6 for a strong acid cation exchanger.

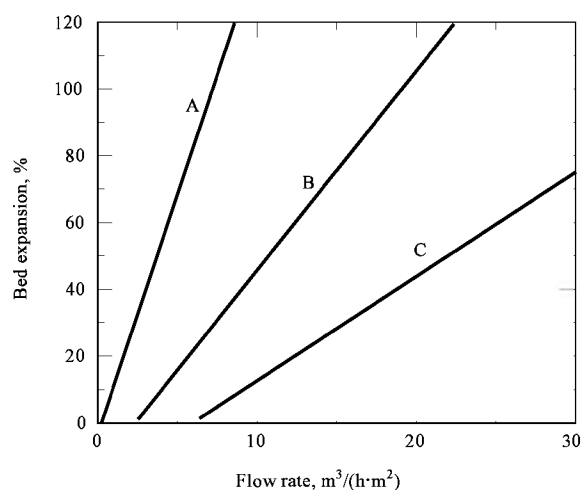


Fig. 6. Bed expansion as affected by resin type, flow rate, and temperature, where A represents a strong base styrenic resin in the Cl⁻ form at 4°C, and B and C a strong acid styrenic resin in the Na⁺ form at 4 and 50°C, respectively (14). To convert m³/(h·m²) to gal/(min·ft²), multiply by 0.409.

Chemical Stability. Oxidants, such as dissolved chlorine [7782-50-5] in water supplies, react with synthetic ion exchangers to cause a loss of capacity, physical weakening of the resin, and partial solubilization of the resin. Anion-exchange resins are most prone to loss of functionality as the oxidant attacks and severs the linkage between nitrogen and carbon on the polymeric structure. In addition to this form of degradation, the functional groups of strong base anion exchangers partially convert to weak base groups through loss of one or more of the alkyl groups attached to the nitrogen. The net effect is loss of both strong base and total capacity with an increase in weak base capacity. Loss of functional groups from cation-exchange resins by oxidative attack is uncommon.

The point at which two polymeric chains are joined together by a cross-linker such as divinylbenzene, or sites where tertiary hydrogens are located in the structure, are other locations for oxidative attack. In both cation- and anion-exchange resins, oxidative attack results in the removal of cross-linking. The moisture content is higher than when the resin was new. Resins having lesser amounts of cross-linkage are more subject to physical deformation when under compressive forces than a resin with a greater amount of cross-linkage. Consequently, a gradual increase in pressure drop across the resin bed is to be expected as oxidative attack continues. In severe cases, the resin breaks into fragments. A resin that has undergone significant oxidative degradation releases small soluble fractions of the polymeric structure to the liquid phase, thus contributing to the biological oxygen demand (BOD). Such groups may or may not interfere with ion-exchange units that follow. An increase in the quantity of water to rinse the column after regeneration is another indicator of oxidative attack, especially for anion-exchange resins. The severance of functional groups leads to the development of carboxylic acid functionality. These groups are converted to the sodium salt when in contact with the NaOH used for regeneration. The water rinse which follows the regeneration step slowly hydrolyzes sodium ions from these groups causing the pH of the effluent to remain high for longer periods of time than realized with new resin.

The rate of oxidative attack is enhanced by the presence of metals such as iron [7439-89-6] and copper [7440-50-8] which serve as catalysts (14), by higher temperatures, and by higher concentrations of oxidants. The direct effect of each of these factors is a matter of opinion because all of the factors causing oxidative degradation are present at the same time but to varying degrees over the useful life of the resin. The tolerable limit for residual chlorine in a water supply is an example. At any concentration, the rate of attack is greater in warmer climates than in colder ones, all other factors being equal. A common recommendation is not to exceed 0.3 mg/L residual chlorine; others may recommend a 0.1 mg/L limit, and still others believe 1 mg/L is a safe upper limit. The oxidant concentration can be lowered before the process stream contacts the resin by standard procedures such as the use of carbon [7440-44-0] columns and the feeding of sulfite [1426-54-5].

Aside from low concentrations of oxidants found in most water supplies, the processing of chemical streams with much higher levels of oxidizing chemicals is practiced occasionally on an industrial basis. The potential dangers are generally recognized. Nevertheless, there is the potential for an uncontrolled reaction that releases heat which converts a liquid to a gas, resulting in the rupture of equipment in an explosive manner. Systems must be designed with appropriate detectors for abnormal performance, and with procedures that reverse or stop the reaction before it gets out of control.

Thermal Stability. Ion-exchange resins should not be used at temperatures above those recommended by the manufacturer. Exceptions are made when frequent replacement of resin is an economic advantage over the operating and capital cost of cooling and reheating the process stream. Functional groups are lost from both cation- and anion-exchange resins when the temperature limit is exceeded (15,16). The rate of loss increases exponentially as the temperature rises above the upper limit. Sulfonated cation exchangers can frequently tolerate temperatures up to 125°C (255°F). Strong base anion exchangers having a trimethyl quaternary ammonium structure can be used up to 77°C (170°F) in salt forms and up to 60°C (140°F) in the OH form. Those strong base resins with a dimethylethanol quaternary ammonium structure and the acrylic anion exchangers are limited to about 40°C (104°F).

Physical Stability. Excessive pressure drop across the resin bed causes fragmentation of the beads. The point at which this occurs depends on the structure of the resin and in most systems is well above the pressure drop listed in product literature for water systems. Upper limits are about 7–9 kg/(cm²·m) (30–40 psi/ft) for gel (microreticular) type resins in the 8–12% DVB range. If oxidative attack occurs during use of the ion-exchange resin the maximum pressure drop would be characteristic of a lower cross-linked resin. Resin breakage aggravates the pressure drop problem. Macroporous resins generally can tolerate somewhat higher pressure drop than gel-type resins (17). Gel-type resins are more resistant than macroporous resins to breakage caused by a shearing motion.

Resins shrink and swell as they are alternately put through adsorption and regeneration cycles. The larger the volume change and the shorter the time involved, the greater the potential for physical damage to the resin particles. In most applications, the greater potential for physical damage occurs during regeneration. However, similar effects occur in a few applications when regenerated resin is contacted with high concentrations of salts, as in the removal of impurities from those salt solutions. The appearance of cracks is the first sign of physical deterioration. Fragmentation into smaller irregularly shaped particles is a sign of further deterioration.

Resins should always be protected from freezing, although that may not always be possible. Generally, a few freeze-thaw cycles do not result in visual damage (cracking or fragmentation). Nevertheless, some weakening of the physical structure occurs because fragmentation is apparent if cycling continues. If operating conditions dictate a lengthy shutdown of the ion-exchange system and the resin columns are in an area that cannot be protected from freezing, the columns may be filled with typical antifreezes without damaging the resin. Neither glycol nor alcohol damage any of the standard cation and anion exchangers. Solutions of NaCl may also be used. When the units are returned to service, they must be thoroughly rinsed with water and, preferably, regenerated before using. The glycol or alcohol must be disposed of in an environmentally approved manner.

Transfer of resin from one vessel to another, as for regeneration, does not physically damage the resin as long as certain practices are avoided. The fluidity of the resin slurry and the linear velocity in transfer lines should be sufficient to keep all particles in suspension. Resin that settles can lead to plugging of the transfer line. Sharp bends in the transfer line should be avoided. If a pump is used for transferring the slurry, it should be a type that will not allow resin particles to be caught in valves when they close. Recessed impeller, peristaltic, and some diaphragm pumps are used in large industrial systems. Recommendations from someone knowledgeable in resin transfer is advised in order to avoid a costly installation that may cause rapid loss of resin through physical damage.

Radiation Stability. Numerous studies have been undertaken to define the effect of radiation on all types of ion-exchange resins. As

expected, the more intense the radiation the greater the damage (18). Cation and anion exchangers lose weight and capacity, cross-linking is removed, and water-soluble components are released if the radiation tolerance limits have been exceeded. The effects of gamma radiation have been studied more than other types of radiation.

Equipment

Ion-exchange systems in process applications may be batch, semicontinuous, or continuous. Batch operations are not common but, where used, involve a kettle with mechanical agitation. Injecting with air or an inert gas is an alternative. A screened siphon or drain valve is required to prevent resin from leaving with the product stream.

Semicontinuous and continuous systems are, with few exceptions, practiced in columns. Most columnar systems are semicontinuous since flow of the stream being processed must be interrupted for regeneration. Columnar installations almost always involve the process stream flowing down through a resin bed. Those that are upflow use a flow rate that either partially fluidizes the bed, or forms a packed bed against an upper porous barrier or distributor for process streams.

The lower section of a column with downward flow must have a distributor system that not only collects liquid evenly over the cross-sectional area, but also supports the resin bed and prevents resin from leaving the column. The traditional method has been to place a network of pipes with small holes drilled in them (a distributor) in a bed of graded gravel, sand, or anthracite coal, which supports the resin bed. While that practice continues, the trend has been toward other approaches. In one modification, the underbed is eliminated by securely wrapping the pipe elements with small mesh, noncorrosive screening. The size of the screen openings must be sufficiently smaller than the resin particles to avoid plugging. Blockage of the openings increases pressure drop and contributes to uneven or channeled flow. Special pipes formed by spirally winding triangular wire around supports, while carefully controlling the space between the flat side of the wire as it is wound, is another approach that is gaining acceptance. Perforated plates separating the resin from the distributor are used in other installations. Careful design of the distributor is essential, especially for the larger diameter units (see FLUID MECHANICS). If the linear flow rate near the wall of the column is substantially less than the midsection of the column, premature breakthrough, more frequent regeneration, and incomplete utilization of the rated operating capacity for the resin result.

The space immediately above the resin bed may or may not be filled with liquid in downward flow systems, depending on the design. If not filled, water entering the column from the top and impinging on the upper surface of the resin bed forms hills and valleys unless the flow is dispersed over the cross-sectional area. A distributor similar to the one used to collect resin below the bed, or splash plate, is placed a short distance above the resin bed to improve the distribution of the process stream flow.

A distributor is frequently installed at the top of the column for use during backwash. It collects water evenly and prevents resin from escaping the column should unexpected surges develop in the water flow during backwash. Columns lacking an upper distributor or screen to prevent loss of resin should have an external system to prevent resin from being lost to the drain. It is referred to as a resin trap and may consist of a porous bag that fits over the outlet pipe or a tank designed to lower the linear velocity. Resin drops to the bottom of the tank and is returned to the column when convenient.

Mixed-bed columns contain an anion and cation exchanger which must be regenerated independently after separation by backwashing. When regeneration is performed in the same column, a distributor is installed near the expected interface of the resins following the backwash. The distributor is used to feed regenerating solutions, feed water, and to collect spent regenerant solutions. Again, distributor design is critical.

All columns, distributors, and ancillary hardware such as piping, valves, and pumps must be constructed of corrosion-resistant materials, or coated with an appropriate substance. All streams that contact the hardware during each step of the cyclic operation need to be considered in this selection.

Columns are designed to have a larger internal volume than the quantity of resin they will contain. The extra space is to provide the necessary volume for a fluidized bed during backwash. Most units are designed for the space above the resin bed (free-board) to be between 50 and 100% of the packed resin bed. Small columns are, on occasion, designed for one-use applications. Since backwashing is of no importance, there is a tendency to fill the unit with as much resin as possible. That practice can be hazardous, especially if the resin swells as a result of oxidative attack or through conversion from one ionic form to another.

Column dimensions vary considerably from one installation to another, depending on the application, total flow, and overall system design. If a tall narrow column and a short wide column contain the same amount of resin and process a stream at the same flow rate, the wide column will have a more favorable linear velocity and a lower pressure drop. However, bed depths cannot be too low, especially in the larger diameter units. Otherwise achieving uniform flow over the cross-sectional area of the column is impossible. The recommended minimum bed depth is about one meter. Bed depths over three meters are most common in applications involving catalysis and chromatographic separations. Serious consideration should be given to the use of several columns in parallel in place of one very large unit, especially when a system cannot be shut down when regeneration is required.

Feed systems utilizing gravity are rarely used. Line pressure is usually adequate for small systems. Auxiliary pumps are required in larger systems to assure proper flow through all units and to avoid uneven flow should line pressure decrease as other demands for water or the process stream occur elsewhere in the facility.

Regenerating streams require dilution with water before contacting the resin. Eductor systems (which mix two fluids), in-line dilution, and separate storage tanks containing sufficient diluted regenerating solutions are commonly used. Appropriate measuring methods to assure that the correct concentration and volume of regenerant are transferred to the resin column should be incorporated in the system. A low concentration, or insufficient volume, adversely affects the subsequent adsorption step. An unintentional higher concentration increases operating costs and magnifies a waste regenerant disposal problem.

Systems

Ion-exchange systems vary from simple one-column units, as used in water softening, to numerous arrays of cation and anion exchangers which are dependent upon the application, quality of effluent required, and design parameters. An illustration of some of these systems, as used in the production of deionized (demineralized) water, is presented in Figure 7.

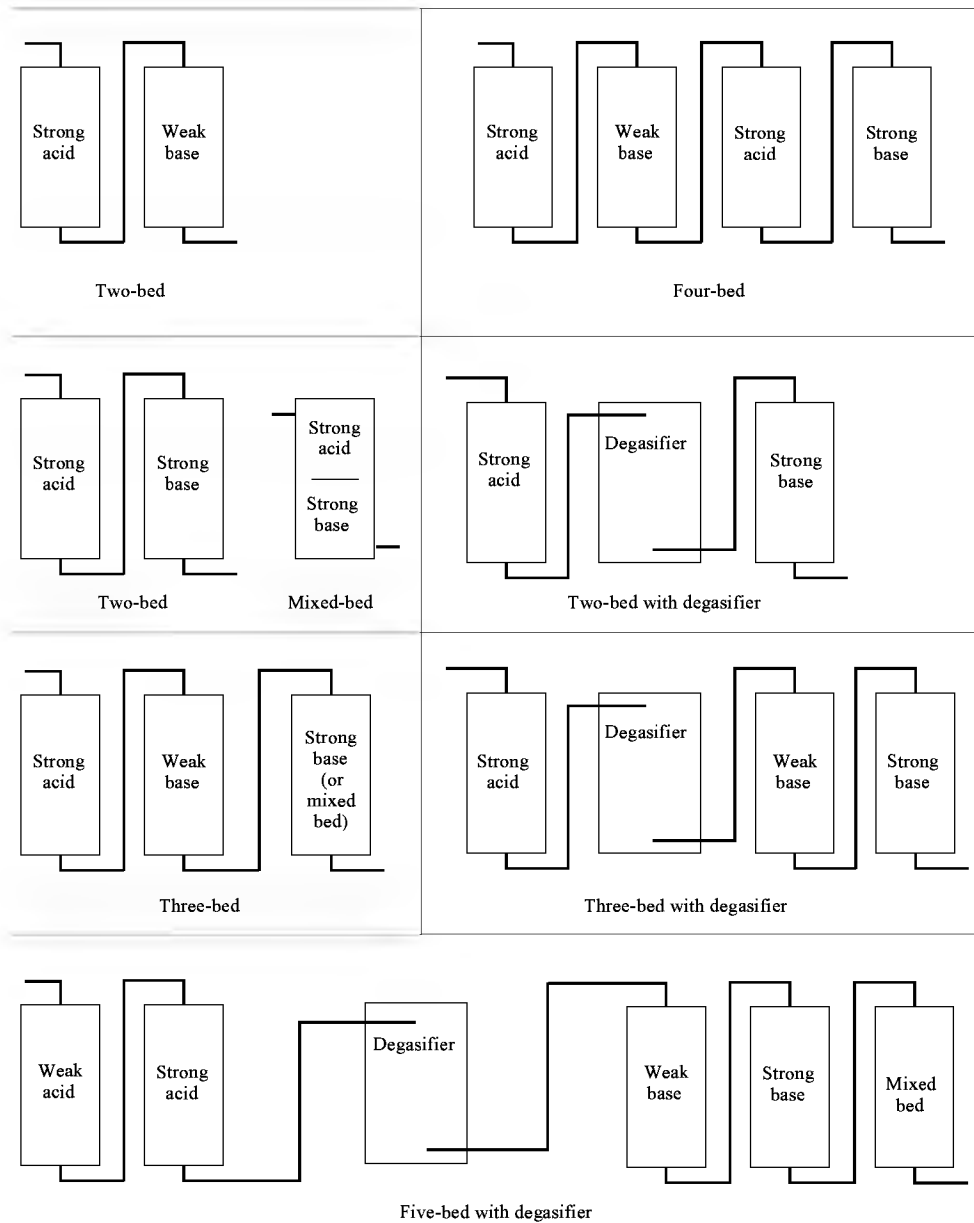


Fig. 7. Various deionization systems. A degasifier facilitates the removal of dissolved gases.

A single-column installation is satisfactory if the unit can be shut down for regeneration. However, if flow of the stream being processed must be continuous, then two or more columns of the same resin must be installed in parallel. Regeneration of each is staged for different time periods.

Two columns, one containing a cation exchanger and one an anion exchanger, are required for a deionization process. The cation exchange unit must be a strong acid-type resin, except in more complex systems, and it must precede the anion-exchange unit. Placing the anion exchanger first (reverse deionization) generally causes problems with precipitates of metal hydroxides, assuming that cations that form these compounds are present in the entering water or process stream. The resin in the anion-exchange unit may be a weak base resin when removal of anions (silicate [12627-13-3], bicarbonate, fluoride, and others) of weak acids is not essential. Otherwise, the column must contain a strong base resin. An alternative is the installation of a third column, which contains the strong base resin, to be placed after the weak base anion-exchanger unit.

A column containing a mixture of cation- and anion-exchange resins is called a mixed bed. Although all types of resin have been considered for these units, the majority consist of strong acid cation and strong base anion exchangers. This system yields a higher quality deionized water or process stream than when the same resins are used in separate columns. Regeneration of a mixed bed is more complicated than of a two-bed system. A mixed bed containing the same total resin volume as in one of the two-bed columns lowers capital costs but must be regenerated at roughly twice the frequency. Mixed-bed units are preferred as final polishers for multiple-bed systems, recycling streams, condensate polishing, and other areas where the electrolyte concentration is in the low mg/L range or less.

Most ion-exchange columns are operated concurrently. Both the process stream and the regenerating solution flow through the resin bed in the same upward or downward direction. Downflow is more common. These streams flow in opposite directions in counterflow systems, as with downflow during the adsorption step and upflow during regeneration. Counterflow provides more efficient use of regenerating chemicals, a higher quality effluent, and higher operating capacity than is obtained with the same resins in concurrent systems. Installation costs are somewhat higher. Extra care must be taken during upflow regeneration since the higher specific gravity of the regenerating stream has a more buoyant effect on the resin particles. Fluidization lessens regeneration efficiency.

Weak acid cation exchangers are used in some deionization systems when bicarbonate alkalinity is higher than in normal water supplies. These resins are regenerated with much greater efficiency than strong acid resins, and operating capacity is higher. The column containing the weak acid resin is installed as the first in a chain or series of columns and removes all or part of the divalent cations, depending on the water composition. In an effort to reduce capital costs for each additional column placed in an ion-exchange system, some designs incorporate a layered bed approach. The two-bed weak acid column followed by a strong acid resin column is reduced to one column by layering the weak acid resin on top of the strong acid resin. A similar anion unit consists of a weak base resin layered on top of a strong base resin. The resins must not be mixed; otherwise, the advantages of using the weaker acidic or basic resin before the stronger version disappears. Much more care must be taken during backwash and regeneration to maintain the layers. Precipitation of calcium sulfate in the cation unit, and precipitation of silica in the anion unit, are potential problems during regeneration.

The most demanding requirements with respect to water quality are in the electronics industry and in very high pressure power plants (see ELECTRONIC MATERIALS; POWER GENERATION). Although mixed-bed units are recognized for giving practically complete removal of all ionic constituents, the mixed-bed unit will give off trace amounts if systems are not designed to approach 100% separation of the two resins before regeneration. Any cation-exchange resin remaining with the anion exchanger is converted to the sodium form when the anion exchanger is regenerated with NaOH. Likewise,

any anion-exchange resin remaining with the cation exchanger is converted to the sulfate form when the cation exchanger is regenerated with sulfuric acid. Resins returned to service, after remixing, gradually release sodium ions or sulfate [14808-79-8] from resin exposed to the wrong regenerating solution. Dissociation of water provides the H^+ and OH^- needed for an exchange of ions to occur. Designs to overcome this problem include addition of an inert resin which occupies space between the anion and cation exchangers after backwash is completed. Special backwash towers have been designed for those systems which do not incorporate the inert resin. For example, a narrower diameter where the interface appears after separation minimizes the volume of resin which might contain some of each resin. That portion is saved for the next regeneration while resin above and below that zone is regenerated in other units.

Numerous efforts have been made to develop continuous ion-exchange systems in which resin moves intermittently or continuously in a direction opposite to the flow of all liquids during adsorption, backwash, regeneration, and rinse. These include resin-in-pulp (RIP) systems used in the uranium industry. Resin is placed in banks of baskets constructed from screens having a mesh size that allows ore particles in an acid or alkaline leached slurry to pass through. The baskets are dipped in and out of a trough through which the slurry passes. The banks of resin are advanced through water washes and regenerations by valving changes for solutions flowing through the troughs (4). In another design, a number of agitated tanks each containing an ion-exchange resin and the leached ore slurry are arranged in series. A vibrating screen is placed between each tank. A mix of resin and ore slurry flowing from a tank was pumped or air lifted to the vibrating screen to separate resin from slurry. Resin moves in one direction to another agitated tank; slurry moves to a different tank in the opposite direction.

Another continuous system consists of columns having numerous perforated plates. Resin enters the top and liquid is pumped into the bottom. The liquid flow rate is adjusted to prevent resin from passing through the openings in either direction. Liquid flow is stopped for a short period to allow resin to drop through openings to the chamber below. Liquid flow resumes before resin can drop through more than one chamber.

A unit referred to as the Higgins Loop has been popular in water treatment, as well as other applications. Resin is pulsed at regular intervals around a rectangularly shaped loop. The diameter of the adsorption section is larger than that of the regeneration section.

A more recent approach, developed by Advanced Separation Technologies (Lakeland, Florida), involves the placement of a number of columns on a carousel that rotates constantly at an adjustable speed. Instead of having one tall fixed column, this system has one taller column which has been broken down into smaller columns on the rotating table. Each is connected in series which allows each column to be on the adsorption cycle beyond the normal breakthrough point typical of the larger column with no other column in series behind it. The number of columns in series during adsorption, backwash, regeneration, and rinse is variable. Liquids flowing into and out of each column change when the column reaches specific positions on the carousel as it rotates.

Many ion-exchange systems incorporate activated carbon [7440-44-0] columns to adsorb dissolved chlorine or high molecular weight organics. Reverse osmosis units are incorporated at times to lower the electrolyte concentration before ion exchange (see REVERSE OSMOSIS). Ultrafiltration (qv) units are installed to lower colloidal solids. Ultraviolet light systems are used to destroy microbiological organisms that tend to propagate in resins and recirculated water systems. Not only do these other water treatment procedures improve the quality of water produced but they also extend resin life.

Lead-lag or merry-go-round systems are more common in other areas of application than in water treatment. In these systems, two or more columns of the same resin are connected in series and the lead column is retained on the adsorption cycle beyond a typical breakthrough point. The lead column is removed for regeneration when the concentration of the effluent from that column is about equal to the influent concentration, or when leakage from one of the succeeding columns reaches a specified limit. The column that was second in line becomes the lead column. The column removed from the circuit is regenerated, rinsed, and then either put back into service in the last position or put on standby. These systems are most useful for recovery of valuable products, or for removal of toxic substances from waste streams. In either case, the objective is to use as much of the resin's total capacity as possible. Adsorbing more valuable product in this manner means a smaller amount of contaminating products are coadsorbed. When the system involves toxic substances, the regenerating stream is at higher concentration for precipitation or other means of disposal.

Cyclic Operation

Resins are seldom used once and discarded. Whether the system is run batchwise or in columns, the resin must be periodically removed from service and regenerated. An exception is the use of a resin as a catalyst in organic reactions. Each cycle consists of two principal steps, adsorption and regeneration, and one or more intermediate steps, rinse and backwash. Failure to use good practices results in poor cyclic performance.

Adsorption. Impurities are removed, or valuable constituents recovered, from a process stream during the adsorption step, which is also referred to as loading or exhausting the resin. Performance is rated primarily on meeting objectives for completeness of removal. Performance is also rated on operating capacity, frequency of regeneration, and operational costs. Variables affecting performance include resin selection, solution chemistry, operating conditions, and equipment design. All are interrelated in varying degrees. Completeness of removal improves by using a resin more selective for that constituent. Using a resin having a selectivity substantially greater than required for the process stream generally results in lower operating capacity, more frequent regenerations, higher operating costs, and higher capital investment. For example, strong acid rather than weak acid cation exchangers are used to soften water supplies. A change in the composition of the stream being processed has a dramatic effect on degree of removal unless regenerating conditions are changed. An increase in the ratio of sodium to divalent cations, or a decrease in the ratio of bicarbonate to sulfate, while maintaining the same total electrolyte concentration, lessens the degree of cation removal in a deionization process when using identical regeneration levels. Equipment must have good distributors for distribution and collection of the process stream flowing through the resin bed. Otherwise, significant quantities of resin have no opportunity to participate in the exchange process (channeled flow). Columns must be designed such that the flow rate through the resin bed is not too fast to accommodate the kinetic properties of the resin, nor too slow.

At times, the process stream flow must be increased after the initial installation to satisfy production demands. Depending on the magnitude of the increase, the existing system may or may not be able to handle the added flow. If it can, regeneration frequency must increase. Adding more resin to the column is often considered an alternative to installing another column. Resin addition lessens the space for backwash and may be a cause for poor column performance.

Degree of removal is usually shown graphically by plotting the effluent concentration as a function of the liquid volume processed. The concentration in the effluent is called leakage, and the plot is referred to as a leakage curve. Very low leakage is common for most ion-exchange installations. It may be below the analytically detectable limit. If above, it may be within acceptable performance requirements. Seeking significantly greater removal than required means higher chemical costs for regeneration because the regenerant level is higher than necessary. The adsorption step is usually terminated when the impurity level in the effluent exceeds a predetermined value, for example, greater than 5 or 10^0 % of the influent concentration. The adsorption step might also be terminated when the effluent conductivity (or resistivity) changes significantly, or when a predetermined time of operation or liquid volume has been reached.

Backwash. Process streams may contain small suspended particles that carry over from a previous processing step. Packed resin beds act as excellent filters for these particles although they are not recommended for this purpose. As suspended particles collect on top of the resin bed, and penetrate deeper down, pressure drop across the bed increases. In severe cases, uniform flow reverts to channeled flow with the liquid seeking the path of least resistance. In other situations, lengthy adsorption runs, though desirable, lead to an increase in pressure drop as the resin particles pack closer together. Both of these conditions are alleviated by backwash. Water is passed up through a bottom distributor at a flow rate sufficient to expand the resin bed by 50–100%, and exits the top of the column. No resin, other than a small amount that may have undergone physical degradation, should escape the unit as long as the column was designed to accommodate that degree of expansion. Tap water is usually of adequate quality for backwashing purposes. Backwash frequency varies from one installation to another. Water availability may be a controlling factor. In applications where the adsorption step exceeds several weeks, the service flow is, at times, interrupted to reverse the direction of liquid flowing through it, as a means to overcome the small degree of packing that took place.

As manufactured, most resins have a Gaussian-like distribution of particle size. Very few are as small as 0.3 mm or as large as 1.0 mm. Most are between 0.5–0.8 mm. A backwash before using new resin is common practice to assure uniform flow during the adsorption and regeneration steps. The backwash eliminates air pockets that may have formed while filling the column and sorts the beads such that the smaller sizes are at the top of the bed and

the larger sizes at the bottom.

Regeneration. The regeneration step is also called elution. The word backflush is used in some areas but leads to confusion with the backwash step and should be avoided. Regeneration is of much shorter duration than the adsorption step. The combined time for backwashing, regeneration, and rinse is usually not longer than two hours. The time is shortened using a smaller volume of regenerating chemicals at a higher concentration, or by increasing the regenerant flow rate. Neither approach should be attempted until experiments have demonstrated that there will be no detrimental effects to overall operation of the ion-exchange system.

Ions have ample time to slowly migrate to the functional groups throughout the resin particles, including those at or near the core, during the adsorption cycle. If adequate time is not provided during regeneration for those same ions to migrate back out of the resin, then two problems, higher leakage and lower operating capacity, are experienced during the next adsorption step. Both conditions can be improved by lowering the regenerant flow rate (increases the contact time), or by using more regenerating chemicals. The latter approach is not only wasteful of regenerating chemicals but also increases waste treatment charges. Whereas the high electrolyte concentration of the regenerating solution provides a strong driving force for removal of ions from the functional groups during regeneration, the slow migration of ions through the resin particles is the main reason for patience in the regeneration step. Flow rates commonly used for regeneration in terms of bed volumes (BV) are between 4–6 BV/h, but 2 BV/h is preferred in many cases. Higher flow rates are used occasionally. However, the advantages are questionable, except in those situations where precipitates might form while the regenerant stream is passing through the resin bed.

A regeneration level that gives 100% conversion of the resin to the regenerated form is economically unsound in almost all applications. The most cost effective regeneration level is the lowest one that will assure a column effluent meeting the required quality during the adsorption step. A higher regeneration level achieves the same goal with a somewhat higher operating capacity. Regeneration levels commonly used in industrial columns are between 1.6 and 3.2 g-equivalents/L resin (4–8 lb/ft³). The lower level of this range is about equal to the total capacity of strong acid cation exchangers (2.0 g-equivalents/L) and strong base anion exchangers (1.4 g-equivalents/L). Complete removal of all ionic constituents is rarely achieved with these resins even when the regeneration level is doubled. With weak acid and weak base resins, on the other hand, there is no advantage in using more than 10 to 15% excess acid or base over the equivalents of ions adsorbed during the previous cycle.

Cation exchangers are regenerated with mineral acids when used in the H⁺ form. Sulfuric acid [8014-95-7] is preferred over hydrochloric acid [7647-01-0], HCl, in many countries because it is less expensive and less corrosive. However, the use of hydrochloric acid is the best method of overcoming precipitation problems in installations which deionize water with high concentrations of barium or calcium compared to other cations. A 4% acid concentration is common, although sulfuric acid regenerations may start as low as 0.8–1% to minimize calcium sulfate [7718-18-9] precipitation. Phosphoric acid [7664-38-2] is rarely used because of cost and disposal problems. Nitric acid [7697-37-2] is to be avoided because it is known to cause catastrophic damage to resin, equipment, and personnel if appropriate controls and monitoring systems are not installed.

Strong base anion exchangers must be regenerated with sodium hydroxide [1310-73-2] when used in the OH⁻ form. Potassium hydroxide [1310-58-3] is a more expensive alternative. Weak base anion exchangers may be regenerated with solutions of ammonium hydroxide [1336-21-6], NH₄OH, or sodium carbonate [497-19-8], Na₂CO₃, although NaOH is more common. The most common concentration for basic regenerating solutions is 4%.

When strong acid cation exchangers are used in the Na⁺ form and strong base anion exchangers are used in the Cl⁻ form, they are regenerated with a 10% sodium chloride [7647-14-5], NaCl, solution. Other concentrations may be used, perhaps with some adjustment in flow rate.

The temperature for regenerating solutions is of little importance except when silica [7631-86-9] removal from a water supply is critical. The less complete the removal of silica from the resin during regeneration, the greater the leakage of it during the subsequent adsorption step. Silica is known to polymerize as it is adsorbed by anion-exchanger resins. As a consequence, depolymerization and solubilization of the silica must be improved by increasing the contact time with NaOH, by increasing the concentration of NaOH, or by increasing the temperature. The last approach is generally used. Temperatures should not exceed 60°C for the Type I resins, and 40°C for Type II and acrylic resins. Thermal degradation and the loss of functional groups occur when these temperatures are exceeded. Elimination of silica from the resin bed is further improved by preheating the bed with warm water before injecting the NaOH solution.

A two-step regeneration is necessary in some applications. One step removes the adsorbed ions, and the second step places the functional groups in another ionic form. For example, weakly acidic cation exchangers are used in the sodium form to remove toxic metals, such as copper [7440-50-8] and nickel [7440-02-0], from neutral or slightly alkaline waste streams. The selectivity of the resin for these metals is so great that NaCl is ineffective for regeneration regardless of the concentration. The accepted route is to elute the metals with a 5–15% acid solution, which leaves the functional groups in the hydrogen form, and then follow with a dilute NaOH solution to convert the groups to the sodium form.

Mixed-bed resins cannot be regenerated until the two resins are separated by backwashing. Each resin is regenerated separately. The cation exchanger should not be in contact with the NaOH solution used for the anion exchanger. The anion exchanger should not be in contact with the acid solution used to regenerate the cation exchanger.

Recycling regenerating chemicals is a practical way to reduce chemical costs. However, the spent regenerating solutions exiting the column should not be recycled directly to the influent end of the column. In the regeneration of a cation exchanger with 4% acid, the acid solution is converted to a salt solution as it passes through the resin bed. The resin is converted to the H⁺ form. If the spent regenerating solution were sent directly back to the column, the salt solution would remove the H⁺ placed on the resin by the acid, and very little would have been gained. A profile of the spent regenerating solution as a function of the volume of solution passed through shows the start of a rapid drop in the salt concentration and a rapid rise in the acid concentration, after a volume of regenerating solution equal to 1.5–1.8 times the volume of the resin bed has passed through the bed. In recycling, collection of the spent regenerating solution should begin at some point after the peak salt concentration has been reached. The volume collected is saved for the next regeneration and is fed to the column prior to using fresh regenerating solutions.

Rinse. When transfer of the required volume of regenerating solution to the column has been completed, a small amount of regenerating solution occupies space immediately above the resin bed, between resin particles in the bed, and within the resin particles. It must be displaced with water before the column can be returned to the adsorption step. Rinsing should begin at the same flow rate as used during regeneration and continue at that rate until a volume of water equal to 1–2 bed volumes has been used. After that, the flow rate is increased to the rate normally used during the adsorption step, and continued at that rate until the effluent is of satisfactory quality, as determined by pH, conductivity, or resistivity. The water need not be at an elevated temperature unless the process stream is above ambient temperature.

Stability. Ion-exchange resins undergo chemical and physical deterioration at varying rates if appropriate care is not taken. Improvements have been made continuously in the physical stability of all synthetic ion exchangers. Since the chemistry of the polymeric structure and of the functional groups have not undergone modification since the introduction of styrenic and acrylic resins, no changes are to be expected in oxidative stability except as to how it relates to the degree of cross-linking.

Shipping

Shipping resins in a water wet condition is standard practice. Removal of water by evaporative methods is expensive and not necessary for the majority of applications since they take place in aqueous systems. Dry resins (almost always strong acid cation exchangers) are required in several catalytic applications in the chemical industry. Water wet resins are packed out by weight and sold by volume. Containers consist of plastic bags and polyethylene-lined budlap bags for small packages in the 28 L (1 ft³) range. Larger containers consist of fiber drums lined with one or more polyethylene bags to prevent wetting and deterioration of the drum. Steel drum shipments are by request. Bulkbag (ca 850–1100 L) shipments have become more common since the early 1980s, especially where unusually large volumes of resin need to be transferred to processing equipment. Overhead space must be available for lifting bulkbags and for bottom discharge into an ion-exchange column or a storage vessel.

Economic Aspects

Commercial producers of synthetic spherical organic ion-exchange resins are listed in Table 2. Estimates for ion-exchange resin production are given from time to time, but must be viewed with a degree of uncertainty. Most sales are through distributors who may or may not be manufacturers of ion-exchange equipment. Purchase prices are determined by competitive bids or the distributors price list which may or may not be lower than the price suggested by the resin manufacturer.

Table 2. Commercial Producers

Company	Country	Trade name(s)
Bayer	Germany	Wofatit, Lewatit
Chemolimpex	Hungary	Varion
Dow	United States, Italy, Germany	Dowex
Mitsubishi Kasei	Japan	Diaion
Ostion	Czechoslovakia	Ostion
Purolite	United States, Wales, Romania	Purolite
Rohm and Haas	United States, France, Japan	Amberlite, Amberlyst, Duolite
Sybron	United States	Ionac

Replacement sales have become an ever increasing percentage of total sales in the mature ion-exchange industry. Economic downturns such as that of the early 1990s affect resin purchases; plans for new installations are abandoned or delayed. Greater effort is made to forestall purchase of replacement resin by directing more attention to operating procedures, or by using chemical methods to restore old resin to a cleaner condition. In addition, quality improvement efforts by resin manufacturers have yielded products having greater physical durability, thereby lessening resin losses caused by physical attrition. Competitive technologies such as reverse osmosis (qv) have had an increasing impact on lowering the volume of resin required for old as well as new installations. Additionally, the once large installed volume of resin in the uranium [7440-61-1] industry has disappeared as environmental concerns over nuclear reactors (qv) have brought a halt to construction and operation of plants of this type, and as the cold war ended (see URANIUM AND URANIUM COMPOUNDS). On the positive side, a substantial growth of installed resin has occurred in processing corn sweetener for use in the beverage industry (see CARBONATED BEVERAGES; SWEETENERS). Significant growth has taken place in catalytic applications, especially for the production of methyl *tert*-butyl ether [1634-04-4] (MTBE), a gasoline octane enhancer used in place of tetraethyllead [78-00-2].

Sales of cation-exchange resins have routinely been slightly more than twice the volume of anion exchangers, although deionization requires more anion exchanger than cation. Anion exchangers also have a shorter life and are replaced more frequently than cation exchangers. These factors are expected to favor greater sales of anion exchangers. One of the reasons for the opposite ratio is the large volume of cation exchangers used in homes and industrial plants for softening water. Another is the chromatographic separation of fructose [57-48-1] and glucose [492-62-6] in the corn sweetener industry which requires large volumes of cation-exchange resin. A similar process for recovery of sucrose [57-50-1] from beet sugar molasses added to the demand for cation exchangers. Catalysis (qv) by ion-exchange resins instead of conventional acids has also contributed substantially to the growth of cation exchangers in the petrochemical industry where resins are used for manufacture of MTBE, the hydrolysis of esters, and the hydration of olefins to form alcohols (see PETROLEUM).

Historically the United States was a primary exporter of ion-exchange resin. As of 1994, the United States imports substantially more than it exports. Because compliance with tightening environmental regulations in the United States impacts on the cost of manufacture, offshore resin is most often lower in price.

Cation exchangers range between \$2100 and \$3900 m³ (\$60–110 ft³). The average price of anion exchangers is about three times higher than that of cation exchangers. Anion-exchange resin production requires more steps than cation exchangers, and the chemicals used in production of anion exchangers are more expensive. The higher the degree of cross-linking for any resin, the greater the manufacturing cost. Macroporous resins command a higher price than microporous resins.

Specifications, Standards, and Storage

Resin manufacturers, equipment suppliers, consultants, and those using ion-exchange resins have worked cooperatively through ASTM to develop a set of standard procedures for measuring critical properties of the resins. Most laboratories use a version of these standards for quality control (qv) purposes, and to learn if any significant changes occur in resin properties after extended use. As more demanding requirements are developing for ultrahigh purity water in the power and electronics industries, additional methodology has been developing by organizations allied with those industries. Ion-exchange resins should be protected from freezing and from temperatures above 60°C. Resins, shipped water wet, should be stored in closed containers. If allowed to dry, they shrink. Rapid rewetting results in rapid swelling to the original water wet volume and contributes to bead breakage.

Health and Safety

Ion-exchange resins are not considered hazardous. However, cation exchangers when in the hydrogen form, and anion exchangers when in the hydroxide form, yield acidic and basic solutions, respectively, when in contact with neutral salt solutions. The corrosive potential should not be overlooked, and skin sensitivity has been reported occasionally, especially when gloves are not used when handling resin. Resins which have been used to remove toxic substances may slowly release these materials if the toxic substances are still attached to the resin. Burning resins, if not incinerated properly, release toxic and odorous fumes. Fumes from burning anion exchangers are particularly foul smelling.

A few resins go through unusually large volume changes when converted from one ionic form to another, when changing from one solvent to another, or when wetting dry resin with water or another solvent. Such changes may cause shattering of glass equipment if constrained. Oxidizing chemicals in contact with ion-exchange resins can result in rapid and uncontrolled degradation which may lead to rupture or bursting of the column in an explosive manner. Moreover, resin beads spilled on a floor are especially dangerous and can be the cause of serious accidents resulting from falls.

Uses

Water Treatment. The two primary applications in water treatment are softening and deionization. Other important but less frequently used applications include dealkalization, softening of produced water, desilicizing, and nitrate removal.

Softening. The use of a cation-exchange resin in the sodium form to remove those ions that cause the water to be hard is called softening. Calcium [7440-70-2] and magnesium [7439-95-4] ions are the prime candidates. Other multivalent cations, such as iron, barium [7440-39-3], and manganese [7439-96-5] are also removed in the process. All are replaced with sodium [7440-23-5]. The softened water prevents problems such as scaling of pipes and heat exchangers. The resin is regenerated with NaCl, usually at a 10% concentration. A more costly alternative is potassium chloride [7440-40-7] when higher levels of sodium ions in the softened water are objectionable.

Problems common to water softening installations consist mostly of fouling by iron and removal of cross-linking from ion-exchange resins. Iron [7439-89-6] is present in water as iron(III) [20074-52-6], iron(II) [15438-31-0], and as colloidal iron oxides, depending on the source of the water. Iron hydroxides precipitate within the resin column when the soluble iron(II) is oxidized to iron(III). The NaCl solution used to regenerate the resin is not effective in dissolving precipitated iron. This results in its accumulation on the resin with each additional cycle. Pretreatment of water by chlorination leaves a residual of dissolved chlorine which, if greater than 0.1–0.3 mg L, begins to oxidatively attack the resin by reducing cross-linking. The rate of cross-linking reduction intensifies as the temperature increases, and when metals such as iron and copper that catalyze oxidative attack are present.

Deionization. The removal of cations and anions from water and replacement of them with hydrogen and hydroxide ions is called deionization. The completeness of the ionic removal is dependent on resin selection, design of the system, operating conditions, and the quality of treated water required. In general, systems become more complex as quality requirements increase.

Use of a strong acid cation exchanger and a weak base anion exchanger in a two-bed system yields the lowest quality water because the weak base resin cannot remove bicarbonate and silicate. That water is entirely satisfactory in numerous installations. If those ions must be removed, other systems such as those shown in Figure 7 produce water to meet requirements. Selecting an appropriate system is not easy and should be reviewed with engineers and consultants well versed in ion-exchange technology.

A problem common to deionization systems involves calcium sulfate precipitation in the cation unit when sulfuric acid is used in the regenerating solution at an excessive concentration or at an unfavorable flow rate. Organic fouling of anion exchangers is another common problem, especially when processing surface water supplies. The decay of leaves, bark, and other vegetation produces large and complex fulvic acid [479-66-3] and humic acids [1415-93-6], which are adsorbed by anion-exchange resins. Only a portion of them are removed during the short time available for regeneration. The result is a cycle to cycle accumulation within the resin particles. The rate of buildup varies from one resin type to another, depending on the structure of the copolymer used for its manufacture.

Dealkalization. An ion-exchange process that lowers the bicarbonate [71-52-3] concentration is termed dealkalization. It is used for those water supplies having a relatively high alkalinity or ratio of bicarbonate to sulfate and chloride also present in the water. Several techniques are available. A lime–zeolite system is a combination of chemical treatment and ion exchange. A lime slurry is added to the water, which may be at an elevated temperature, to precipitate most of the calcium and magnesium. A column containing a resin appropriate for softening water removes the residual calcium and magnesium. Systems running close to the boiling point of water are more aggressive with respect to the physical stability of the resin, especially when ambient temperature brine is used for regeneration of the resin.

Another dealkalization system is referred to as a split-stream process. A portion of the incoming water passes through a strong acid cation-exchange resin in the hydrogen form. All cations are removed resulting in an acidic effluent. The other portion of the influent is passed through a parallel column containing a strong acid cation exchanger in the sodium form. This column softens the water by removing calcium and magnesium, but has no effect on the alkalinity. The effluents from both columns are combined to form a neutral stream with much lower bicarbonate content.

Another alternative involves the use of a weak acid cation exchanger in the hydrogen form. This resin is not capable of removing all cations. It removes only the amount equivalent to the bicarbonate in the influent water. The acidity in the effluent stream is carbonic acid [463-79-6] which can be eliminated by installing a degasifier.

Anion-exchange resins are also capable of lowering alkalinity. A Type II strong base resin is preferred and is used in the chloride or mixed chloride–hydroxide form. Regeneration with a NaCl solution containing NaOH gives higher operating capacities than NaCl alone. A problem common to dealkalization is low operating capacity. This results in frequent regenerations.

Produced-Water Softening. In secondary-oil recovery projects involving steam injection to heat oil remaining in strata and to make it more fluid, the steam condenses and the water becomes contaminated with calcium, magnesium, and other salts (see PETROLEUM). This water is cycled back to steam generators after it is separated from the oil. However, severe scaling results if the water is not softened before the generator. Softening is a challenge because NaCl concentrations are usually in the several thousand mg L range, or higher. The greater the Na^+ concentration with respect to Ca^{2+} and Mg^{2+} , the lower the degree of softening and the lower the operating capacity. At the lower total salt concentrations, two columns of strong acid resins are used in series. The second column acts as a polisher and is regenerated with an NaCl solution in an upflow situation. At higher total salt concentrations, a weak acid resin, which has greater selectivity for divalent cations, is used in place of the strong acid resin. At very high total salt concentrations, a chelate resin is used in place of the weak acid resin.

A problem common to produced water applications is the tendency for oil fouling of the resin. If weak acid or chelate resins are used, a two-step regeneration process is required which uses acid to remove calcium and magnesium from the resin, followed by a dilute NaOH solution to convert the resin to the sodium form.

Nitrate Removal. In those areas where nitrate [14797-55-8], NO_3^- , concentrations in the water supply have been close to or above allowable limits for potable water, nitrate removal is practiced. Strong base anion-exchange resins are used in the chloride form. The effluent must be carefully monitored since selectivity favors sulfate over nitrate. Processing water beyond an acceptable nitrate leakage can result in the nitrate concentration exceeding the influent concentration. Newer strong base resins functionalized with triethylamine, tributylamine, and other amines overcome this problem because the selectivity for nitrate is greater than sulfate (19,20). The resin is regenerated with a sodium chloride solution.

Condensate Polishing. Steam is recovered as condensate after passing through turbines, a process commonplace in the electrical power industry. During the process, small amounts of soluble and insoluble impurities appear in the condensate. Recycling the hot water, after removing impurities with mixed-bed resins, is far more economical than treating cold water sources. The rate of flow through the mixed-bed resin is much higher than in conventional water treatment because of the very low level of impurities. Two systems most frequently used are columns (deep beds) and thin layers of ground resin deposited on a pre-coat filter (powdered system). The process is called condensate polishing.

Boron Removal. Boron [7440-42-8] is occasionally present in water supplies at an unacceptable level. It cannot be removed with the standard anion-exchange resins unless the water is deionized. Selective removal is possible by using an anion exchanger functionalized with *N*-methylglucamine [6284-40-8]. This resin is in limited commercial supply. The borate form of conventional strong base anion exchangers is used in some nuclear reactors to adjust the concentration of boron in water used as a moderator. The resin releases boron as the water temperature rises.

Food Processing. The sugar and corn sweetener industries have the largest volume of installed ion-exchange resin in the food processing (qv) industry. Lesser quantities are used to process wine (qv), whey, fruit juices (qv), and gelatin (qv).

Sugar, or sucrose [57-50-1], is obtained from sugar cane as a juice by pressing cut canes, and from sugar beets by slicing the beets and extracting the sucrose with hot water. Organic and inorganic impurities must be removed from these extracts to obtain a white, crystalline product (see SUGAR).

Cane sugar factories produce a raw crystalline sucrose which is shipped to the refinery where the raw sugar is redissolved and impurities are removed by precipitation, bone char or granular carbon, ion exchange, and crystallization (qv). Resins are used for several purposes. Organic compounds responsible for color in sugar syrups are removed by char or carbon in most refineries. Additional removal is achieved by following the char or carbon system with columns of a strong base anion exchanger in the chloride form (21–23). There is a trend toward replacing char and carbon systems with resin for decolorization purposes. Deionization with a weak acid cation exchanger and a strong base anion exchanger is incorporated at those refineries desiring to produce a sucrose syrup with low color and a low ash content. A weak acid resin is selected over a strong acid resin to minimize conversion of sucrose to invert sugar (glucose and fructose). The soft drink industry, on the other hand, prefers invert sugar over sucrose as a syrup for sweetening purposes. Sucrose is inverted, or converted, to an approximate 50–50 mix of fructose and glucose by hydrochloric acid or a strong acid cation exchanger in the hydrogen form.

Beet sugar factories do not produce an intermediate raw, crystalline sucrose. Instead, the thin juice obtained from the beets passes directly through purification processes similar to those used by the cane sugar refineries. Ion exchange, as a processing step in beet sugar purification, is more common outside the United States. The U.S. beet sugar industry has, in recent years, made a more concerted effort to recover sucrose lost to molasses. Residual calcium and magnesium present in the molasses is first removed with a weak acid cation exchanger in the sodium form. The molasses is next passed through a strong acid cation exchanger in the sodium form. The functional groups of this resin slow down the forward movement of sucrose, but not sodium and potassium. The sucrose concentration builds up in zones as the molasses passes through the column. Water is fed alternately with molasses to aid in developing the separation. The hydrogen form of the resin cannot be used to remove sodium and potassium because the resulting acidity would convert a large portion of sucrose to a mixture of glucose and fructose.

Syrup derived from corn starch is called corn sweetener. Starch (qv) is converted to glucose enzymatically or by acid hydrolysis. Color, color precursors, and salts are removed by ion exchange. Most systems consist of several pairs of a strong acid cation exchanger followed by a weak base anion exchanger connected in series. The syrup, thus purified, is used in numerous food products. This same purified syrup is enzymatically processed to convert a portion of the glucose to fructose. In a process similar to the recovery of sucrose from molasses, as mentioned above, glucose and fructose are separated chromatographically by using a strong acid resin. In this application, however, the resin is used in the calcium form. Glucose- and fructose-rich streams are

recovered separately in the effluent and blended to form a high fructose corn sweetener (HFCS) which is of value to the beverage industry. It is similar to the invert syrup derived from sucrose.

Whey, a by-product in the manufacture of cheese, is deacidified and deionized using a two-bed system consisting of a strong acid cation exchanger followed by a weak base anion exchanger (24,25). A mixed bed of a strong acid cation and a strong base anion exchanger is included, at times, after the two-bed system to achieve a higher degree of purification.

Fruit juices can be deacidified with a weak base anion-exchange resin. Removal of compounds which cause a bitter taste is a more popular application (26,27). It is accomplished with resins that have no ion-exchange functionality. In essence, they are similar to the copolymer intermediates used by resin manufacturers in the production of macroporous cation and anion exchangers. These products are called polymeric adsorbents. They are excellent for removal of limonin [1180-71-8] and naringin [10236-47-2], the principal compounds responsible for bitterness in orange, lemon, and grapefruit juices. The adsorbents are regenerated with steam or alcohol. Decaffeination of coffee (qv) and tea (qv) is practiced with the same polymeric adsorbents (28).

Wines are processed by ion exchange for two purposes: excess acidity, which is responsible for tartness, is removed using a weak base anion exchanger (29,30); newly fermented wine is supersaturated with respect to potassium bitartrate [868-14-4] and, unless the concentration is reduced, a precipitate eventually forms. Precipitation is hastened in the traditional method of processing wine by storing it at a lower temperature (chill-proofing). The sediment is periodically removed (racking) (31). The chill-proofing process is substituted with an ion-exchange process at numerous wineries. Precipitation is prevented by converting a portion of the potassium bitartrate to the more soluble sodium bitartrate [526-94-3] when passing the wine through a strong acid cation exchanger in the sodium form.

Pharmaceutical. Ion-exchange resins are useful in both the production of pharmaceuticals (qv) and the oral administration of medicine (32). Antibiotics (qv), such as streptomycin [57-92-1], neomycin [1404-04-2] (33), and cephalosporin C [61-24-5] (34), which are produced by fermentation, are recovered, concentrated, and purified by adsorption on ion-exchange resins, or polymeric adsorbents. Impurities are removed from other types of pharmaceutical products in a similar manner. Resins serve as catalysts in the manufacture of intermediate chemicals.

Ground ion-exchange resins have been used for many years as carriers for drugs which are ingested (see DRUG DELIVERY SYSTEMS). This method of dosing overcomes objectionable odors and tastes. Resins, especially those having strong acid or base functionality, provide slow (or sustained) release over many hours for those medicines adsorbed by them. A low cross-linked modification of a strong base styrene–divinylbenzene resin is dried and ground, then ingested for adsorption of bile acids in the treatment of people having high levels of blood cholesterol [57-88-5]. The pharmaceutical generic name for this resin is cholestyramine [11041-12-6]. A hemoperfusion system incorporating a polymeric adsorbent was developed to adsorb drug and drug metabolites from the blood of patients who had overdosed (see CONTROLLED RELEASE TECHNOLOGY, PHARMACEUTICAL).

Catalysis. Ion-exchange resins, especially the strong acid type, have long been recognized as excellent substitutes for sulfuric acid and other similar catalytic agents. Resins participate in fewer side reactions, and because they are insoluble, remain in the reactor and do not contribute to downstream corrosion problems. Neutralization of the reaction product is not necessary. Regeneration of a resin catalyst is not required unless the incoming reactant streams contain impurities that would be adsorbed by the resin. Separate ion-exchange units are used, at times, to remove the impurities from those streams before they enter the reactor. Numerous large reactor columns have been designed and installed in the petrochemical industry which have no facilities for regeneration. Resin is replaced when performance drops off, which may be well over one to two years of continuing operation.

An etherification reaction to form methyl *tert*-butyl ether [1634-04-4] (MTBE), CH₃OC(CH₃)₃, used as an octane enhancer in gasoline, has become the most widely used application in catalysis by ion-exchange resins and incorporates the largest sized reactors (35,36). The entering methanol and isobutylene [115-11-7] streams are essentially free of impurities. Mole ratios of the reactants and operating temperatures are site specific. The reaction temperature is gradually increased as the resin ages. The preferred catalyst for this reaction is a strong acid-type resin having macroporous properties. The same type catalyst is used in the hydration of olefins to form alcohols, the alkylation of phenols, and the hydrolysis of esters. Microporous catalysts are preferred for the manufacture of bisphenol A [80-05-7], esterification reactions, and in alkylation of phenol [108-95-2]. Macroporous catalysts are also used in several of these processes. At times, it is advantageous to convert the functional groups to other forms, as in the adsorption of platinum [7440-06-4] and palladium [7440-05-3], for example.

Chemical Purification. Many organic and inorganic products manufactured in commercial quantities contain objectionable impurities which can be removed by ion exchange. Selection of the appropriate resin is important. Iron(III) is removed from hydrochloric acid using a strong base anion exchanger (see HYDROGEN CHLORIDE). Divalent cations, eg, calcium, magnesium, strontium, and barium, are removed from saturated NaCl solutions or other monovalent salt solutions using chelate resins (37). Formic acid is removed from hot, concentrated formaldehyde [50-00-0] using a weak base anion exchanger. Amines are removed from methanol [67-56-1] using a strong acid-type cation exchanger. Salts are removed from dimethylformamide [68-12-2] with a strong acid cation exchanger followed by a weak base anion exchanger. Gelatin (qv) [9000-70-8] is purified in a similar manner, although a strong base anion exchanger is generally preferred over a weak base resin. Oxazole [288-42-6] is removed from acrylonitrile (qv) [107-13-1] using a strong acid resin.

Ion exclusion and ion retardation are single-resin systems which separate electrolytes from nonelectrolytes without the normal type of adsorption occurring. Ion-exchange resins exclude a fraction of electrolyte from the resin phase because of the Donnan membrane equilibrium principle. This principle requires equal electrochemical potentials for the permeating ions on each side of the membrane. A column of cation- or anion-exchange resin, with the functional groups in the same ionic form as the cations or anions in a feed stream containing an electrolyte and a nonelectrolyte, excludes most of the electrolyte and slows forward movement of the nonelectrolyte. Separation is achieved by alternating feed of water with the electrolyte or nonelectrolyte stream. Regeneration is not required.

Ion retardation is the opposite of ion exclusion. Instead of using a resin with all sites having the same structure and being in the same ionic form, ion retardation involves a resin which contains both cation and anion functionality. These resins are not widely available commercially. The functional sites are in very close proximity of each other, and partially neutralize the positive and negative charges of the groups. Cations and anions, present in a feed stream containing electrolytes and nonelectrolytes, are loosely held by the functional groups. They are retarded in their forward motion, while the nonelectrolyte passes through the resin bed at a slightly faster rate. Like ion exclusion, water is fed alternately with the feed stream, and regeneration is not required. In both systems, water displaces the nonelectrolyte, or electrolyte, that was restrained in its forward movement by the resin.

Metal Processing. Plating, etching, anodizing, pickling, and galvanizing involve chemical solutions or baths that are used repeatedly until the impurity concentration increases to a level where additional use of the bath impairs performance. Destruction of the bath is costly and creates a disposal problem. Ion-exchange units are installed at numerous locations to lower the impurity concentration to an acceptable value. Complete removal is not necessary. Examples include the removal of iron, copper, and trivalent chromium(III) from chromic acid [1333-82-0] plating baths (38), the removal of copper from etching solutions (39,40), the removal of aluminum [7429-90-5] from anodizing baths (41), removal of iron from pickling acids, and removal of iron from acidic zinc sulfate [7733-02-0] ZnSO₄, galvanizing baths (42). Resin selection is critical for success of the process (see ELECTROPLATING; METAL TREATMENTS).

Hydrometallurgy. Uranium [7440-61-1] recovery from sulfuric acid leaching (43–46) and bicarbonate leaching (47,48) operations involved the largest use of ion-exchange resins in the hydrometallurgical area (see METALLURGY, EXTRACTIVE). Activity was at its peak in the 1970s. Many of the mill sites shut down when the demand for uranium in the arms race and in nuclear reactors for generation of power declined. The volume of resin installed in the United States, Canada, and the Republic of South Africa is not likely to be matched in any hydrometallurgical application in the future.

Uranium ores are leached with dilute sulfuric acid or an alkaline carbonate [3812-32-6] solution. Hexavalent uranium forms anionic complexes, such as uranyl sulfate [56959-61-6], UO₂(SO₄)²⁻, which are more selectively adsorbed by strong base anion exchangers than are other anions in the leach liquors. Sulfate complexes are eluted with an acidified NaCl or ammonium nitrate [6484-52-2], NH₄NO₃, solution. Carbonate complexes are eluted with a neutral brine solution. Uranium is precipitated from the eluent and shipped to other locations for enrichment. Columnar recovery systems were popular in South Africa and Canada. Continuous resin-in-pulp (RIP) systems gained popularity in the United States since they eliminated a difficult and costly ore particle leach liquor separation step.

Other hydrometallurgical uses for resin have been small in comparison. Replacement of carbon as an adsorbent for gold [7440-57-5] from a cyanide [57-12-5] leached ore has been studied for many years, but remains in limited commercial use. A deterrent has been the failure to develop an efficient and

safe method to recover gold from the resin. Gold forms an anionic cyanide complex which is readily adsorbed by a strong base anion exchanger (49,50). A program has been underway in South Africa to use weak base anion exchangers (phenolics have shown greatest promise), but since these resins are more sensitive to pH above 7.0, loading capacity is not as high as with strong base resin or carbon. Strong acid resins were installed for the commercial separation of rare earths (see LANTHANIDES). Nickel [7940-02-0] and cobalt [7440-48-4] were separated and purified with an iminodiacetic acid [142-73-4] type chelate resin. Niobium [7440-03-1] has been separated from other metals by forming anionic fluoride complexes and passing the solution through a strong base anion exchanger. Bromine [7726-95-6] and iodine [7553-56-2] are recovered from seawater and other brines by adsorption of bromate [15541-45-4] or iodate [15454-31-6] on a strong base resin.

Waste Treatment. Environmental concerns have increased the need to treat liquid discharges from all types of industrial processes, as well as runoffs where toxic substances appear as a result of leaks or following solubilization (see WASTES, INDUSTRIAL). One method of treatment consists of an ion-exchange system to remove the objectionable components only. Another involves complete or partial elimination of liquid discharges by recycling streams within the plant. This method is unacceptable unless a cyclic increase in the impurities is eliminated by removing all constituents prior to recycling.

In the first approach, numerous toxic metals are lowered below discharge limits through use of the standard cation- and anion-exchange resins, or by using specially formulated resins which have much higher selectivities. For example, chromium(III) [16065-83-1] in a rinsewater or a groundwater is adsorbed on a strong acid cation exchanger (51). Chromium(VI) [18540-29-9] is adsorbed on a strongly basic anion exchanger. Copper is adsorbed more completely from an ammonium sulfate [7783-20-2] solution by chelate resins than by the standard cation exchangers. A resin containing thiol (SH) functionality has very high selectivity for mercury [7439-97-6], and other metals which form sulfide precipitates (52,53). Selective removal is preferred for leachates common to existing and abandoned mine sites, holding ponds, land fills, and lagoons.

Recycling systems incorporate cation and anion exchangers to remove all electrolytes. Other nontoxic material, common to the process, appear in the process water along with the toxic substances and, unless removed, lower the quality of the water. By using a cation- and anion-exchange resin system, the water produced will be of higher quality than water normally available to the facility. Deionization of recycled water is less costly if the electrolyte concentration is lower than in external water supplies. If the concentration is greater in recycled water, elimination of streams with high electrolyte concentrations and free of toxic substances should be considered. Recycling lowers regeneration chemical costs and minimizes the volume of water that must be purchased from an external source.

Numerous organic compounds cannot be removed efficiently with ion-exchange resins or polymeric adsorbents, but others can. Low molecular weight organic acids and bases are adsorbed by anion and cation exchangers, respectively. Whereas higher molecular weight compounds might be adsorbed satisfactorily, they do not desorb easily during regeneration. Polymeric adsorbents are excellent materials for the removal of phenolic-type compounds (54).

Removal of radioactive ions is accomplished with standard resins when selectivities are favorable and when the presence of other electrolytes does not interfere. Deionization systems are common when completeness of removal is essential.

Toxic substances adsorbed on resins are removed during a regeneration procedure. The resulting spent regeneration solution has a higher concentration of the toxic substance than the stream from which it was removed by the resin. Toxic material in the spent regenerating solution can usually be precipitated, electrodeposited as in an electrolytic cell, or made insoluble by other acceptable procedures.

Gas Adsorption. There are few commercial installations. Ammonia [7664-41-7] is adsorbed by a cation exchanger in the hydrogen form and eluted with an acid to give ammonium sulfate or ammonium chloride [12125-02-9]. Success has been reported on the removal of sulfur dioxide [7446-09-5] on a weak base anion exchanger (55) (see ADSORPTION, GAS SEPARATION). Chemical compounds such as phenol, ethylene dichloride [107-06-2], and benzene [71-43-2] have been successfully removed on polymeric adsorbents (56). The concern with systems for removing impurities from air, or other gaseous streams, is the high pressure drop typical of high velocities through beds of small-diameter resin particles. Other concerns are water content of both the resin and the gaseous stream, temperature, and cost effective regeneration procedures, especially for organic substances.

Analytical. Ion-exchange resins have been extremely valuable for a variety of analyses. Total ion electrolyte concentration can be determined by analyzing for total cations or total anions using a cation exchanger in the hydrogen form or an anion exchanger in the hydroxide form. Ions present in solution at very low concentrations are concentrated by adsorption on a resin before eluting and analyzing the effluent by standard procedures. Ions that are interferences for analytical procedures are eliminated by adsorption on an ion exchanger. The progress of a large-scale plant reaction can be monitored by following the disappearance of a reactant. Impurity levels in a finished product are determined by ion exchange if it is adsorbable. Numerous commercial processes have evolved from analytical separations and purifications practiced in the laboratory as an analytical procedure.

Ion chromatography (ic) is a highly valued and growing methodology for analytical analysis of ionic constituents in aqueous streams. In contrast to the chromatographic separations mentioned earlier with conventional resins, ion chromatography uses similar, yet different, resins which yield separations that are measured in minutes, rather than hours and days (see CHROMATOGRAPHY). Resins are somewhat smaller in size and have most of the functional groups on or near the outer surface, in contrast to being distributed throughout the resin matrix in conventional resins. The outer surface functionality shortens the path from liquid phase to resin phase and is the factor not only for more rapid separations, but also separations with little overlap in peaks for separate ions. With functionalization limited to the outer shell, the capacity of the resin is significantly reduced. Ion chromatography is generally considered for the more dilute streams where concentrations extend down to the mg/L and µg/L ranges. Ion chromatography resins are placed in a narrow separator column which is followed by a suppressor column and an analytical instrument to pick up signals in the effluent stream. Separate systems are used for cations and anions.

BIBLIOGRAPHY

"Ion Exchange" in *ECT* 1st ed., Vol. 8, pp. 1–17, by R. Kunin, Rohm & Haas Co.; in *ECT* 2nd ed., Vol. 11, pp. 871–899, by R. M. Wheaton and A. H. Seamster, The Dow Chemical Co.; in *ECT* 3rd ed., Vol. 13, pp. 678–705, by R. M. Wheaton and L. J. Lefevre, Dow Chemical U.S.A.

1. F. Helfferich, *Ion Exchange*, McGraw-Hill Book Co., Inc., New York, 1962.
2. F. C. Nachod and J. Schubert, eds., *Ion Exchange Technology*, Academic Press, Inc., New York, 1956.
3. C. Calmon and T. R. E. Kressman, eds., *Ion Exchange in Organic and Biochemistry*, Interscience Publishers, New York, 1957.
4. R. Kunin, *Ion Exchange Resins*, Robert E. Krieger Publishing Co., Huntington, N.J., 1972.
5. J. A. Kitchener, *Ion Exchange Resins*, John Wiley & Sons, Inc., New York, 1957.
6. J. E. Salmon and D. K. Hale, *Ion Exchange: A Laboratory Manual*, Academic Press, Inc., New York, 1959.
7. K. Dorfner, *Ion Exchangers*, Ann Arbor Science Publishers, Ann Arbor, Mich., 1973.
8. R. T. Hutchinson and J. Katsahian, *The Manufacture of Liquid Sugars*, Rohm and Haas, Philadelphia, 1977.
9. C. Calmon, *React. Polym.* **4**(2), 131–146 (1986).
10. **U.S. Pat. 2,366,007** (Dec. 26, 1945), G. F. D'Alelio (to General Electric Co.).
11. **U.S. Pat. 2,596,417** (Apr. 22, 1952), G. F. D'Alelio (to General Electric Co.).
12. N. W. Frisch, F. McGarvey, R. Kunin, and J. W. Moffett, *Proceedings of the 16th Annual Water Conference*, Engineers Society of Western Pennsylvania, Oct. 1955.
13. S. Peterson, *Ann. N. Y. Acad. Sci.* **57**, 144–158 (1953).
14. *Amber-hi-lites*, No. 161, Rohm and Haas Co., Philadelphia, Spring 1979.
15. L. Petrus, E. J. Stamhuis, and G. E. H. Joosten, *Ind. Eng. Chem. Prod. Res. Dev.* **20**(2), 366–371 (1981).
16. P. E. Tulupov, T. Y. Butenko, and E. F. Radaev, *Zh. Prikl. Khim. (Leningrad)* **54**(10), 2218–2221 (1981).
17. M. C. Skriba, W. M. Alvino, and R. Kunin, *Ind. Eng. Chem. Proc. Res. Dev.* **20**(2), 391–395 (1981).
18. M. Kubota, Y. Konami, H. Nakamura, and H. Amano, *J. Radioanal. Chem.* **45**(1), 73–89 (1978).
19. **U.S. Pat. 4,479,877** (Oct. 30, 1984), G. A. Guter (to U.S. Environmental Protection Agency).
20. T. Tarjon *Polytech., Chem. Eng.* **33**(3–4), 287–298 (1989).

21. C. Liker, *Int. Sugar J.* **86**(1030), 260–265 (1984).

22. L. Ramm-Schmidt, E. Talvitie, M. Tylle, and E. Looma, *Int. Sugar J.* **91**(1084), 65–70 (1989).

23. L. S. M. Bento, *Proc. Sugar Process. Res. Conf. 1990*, 99–115 (1991).

24. R. Delbeke, *Neth. Milk Dairy J.* **33**(4), 181–192 (1979).

25. **U.S. Pat. 5,084,285** (Jan. 28, 1992), M. Shimatani and co-workers (to Snow Brand Milk Products Co., Ltd.).

26. R. Couture and R. Rouseff, *J. Food Sci.* **57**(2), 380–384 (1992).

27. C. W. Wilson, III, C. J. Wagner, Jr., and P. E. Shaw, *J. Agric. Food Chem.* **37**, 14–18 (1989).

28. **Brit. Pat. 1,488,340** (Oct. 12, 1977), (to General Foods Corp.).

29. W. R. Bonorden, C. W. Nagel, and J. R. Powers, *Am. J. Enol. Vitic.* **37**(2), 143–148 (1986).

30. **U.S. Pat. 4,156,026** (May 22, 1979), E. V. Gogel (to Mogen David Wine Corp.).

31. R. W. Percival, F. X. McGarvey, and H. O. Sonneman, *J. A.O.A.C.* **41**(1), 144–151 (1958).

32. O. L. Sprockel and J. C. Price, *Drug Dev. Ind. Pharm.* **16**(2), 361–376 (1990).

33. D. Y. Cha, *React. Polym., Ion Exch., Sorbents* **5**(3), 269–279 (1987).

34. Czech. CS 203,352 (Sept. 15, 1982), E. Lacinova, R. Frimm, L. Welward, and R. Kosalko.

35. **U.S. Pat. 4,316,724** (Feb. 23, 1982), S. Herbstman (to Texaco Inc.).

36. **U.S. Pat. 4,978,807** (Dec. 18, 1990), L. A. Smith (to Chemical Research & Licensing Co.).

37. **U.S. Pat. 5,104,500** (Apr. 14, 1992), W. W. Ruthel (to Occidental Chemical Corp.).

38. B. A. Bolto and L. Pawlowski, *Effl. Water Treat. J.* 6–17 (Jan. 1983).

39. **U.S. Pat. 4,269,716** (May 26, 1981), M. I. Gurian (to Advanced Systems, Inc.).

40. K. F. Keating and M. A. Gouch, *Plating Surface Finish*. 106–109 (Aug. 1986).

41. **Ger. Pat. 139,872** (Jan. 23, 1980), U. Fisher.

42. **Rus. Pat. 79,161,547** (Dec. 21, 1979), T. Nagashima, K. Tsuchiya, and M. Mazuma (to Nippon Steel).

43. D. E. Traut, I. L. Nichols, and D. C. Seidel, *U.S. Bureau of Mines Report Investigation*, RI 8282, 1978, Washington, D.C., 29 pp.

44. M. Streat and G. N. J. Takel, *J. Inorg. Nucl. Chem.* **43**(4), 807–814 (1981).

45. F. X. McGarvey and J. Ungar, *J. S. Afr. Inst. Min. Metall.* **81**(4), 93–100 (1981).

46. W. I. Harris, L. B. Lindy, and R. S. Dixit, *React. Polym., Ion Exch., Sorbents* **4**(2), 99–112 (1986).

47. D. E. Traut, N. M. T. El Hazek, G. R. Palmer, and I. L. Nichols, *U.S. Bureau of Mines, Report Investigation* RI 8328, 1979, Washington, D.C., 13 pp.

48. **U.S. Pat. 4,312,838** (Jan. 26, 1982), T. Y. Yan (to Mobil Oil Corp.).

49. F. E. Ahmed, B. D. Young, and A. W. Bryson, *Hydrometallurgy* **30**(1–3), 257–275 (1992).

50. J. B. Hiskey and V. P. Atluri, *Miner. Metall. Process* **6**(3), 112–116 (1989).

51. A. K. Sengupta and D. Clifford, *Environ. Sci. Technol.* **20**(2), 145–155 (1986).

52. A. B. Osteen and J. P. Bibler, *Water, Air, Soil Pollut.* **56**, 63–74 (1991).

53. J. A. Ritter and J. P. Bibler, *Water Sci. Technol.* **25**(3), 165–172 (1992).

54. C. R. Fox, *CEP*, 70–77 (Aug. 1979).

55. L. Layton and G. R. Youngquist, *I&EC Proc. Des. Dev.* **8**(3), 317–324 (1969).

56. Y-C. Sun and G. R. Killat, *Hydrocarbon Proc.* 241–242 (Sept. 1976).

Charles Dickert
Consultant

ION IMPLANTATION

Modern technology depends on materials with precisely controlled properties. Ion beams are a favored method (and in integrated circuit technology, the prime method) to achieve controlled modification of surfaces and near-surface regions. In every integrated circuit production line there are ion implantation systems. In addition to integrated circuit technology, ion beams are used to modify the mechanical, tribological, and chemical properties of metals, intermetallics, and ceramics without altering their bulk properties.

Ion implantation of materials results from the introduction of atoms into the surface layer of a solid substrate by bombardment of the solid with ions in the eV to MeV energy range. Several ballistic-like atomic processes occur during ion implantation. The ballistic interactions of an energetic ion with a solid are shown schematically in Figure 1. The figure shows sputtering events at the surface, single-ion single-atom recoil events, the development of a collision cascade involving a large number of displaced atoms, and the final position of the incident ion. The solid-state aspects of ion implanted materials are particularly broad because of the range of physical properties that are sensitive to the presence of trace amounts of foreign atoms. Mechanical, chemical, electrical, optical, magnetic, and superconducting properties are all affected and indeed may even be dominated by the presence of such foreign atoms. The use of energetic ions affords the possibility of introducing a wide range of atomic species independent of thermodynamic factors, thereby making it possible to obtain impurity concentrations and distributions of particular interest. In many cases, these distributions would not be otherwise attainable. Interest in ion implantation has been stimulated by the possibilities of synthesizing novel materials with potential applications in the semiconductor, tribological, corrosion, and optical fields.

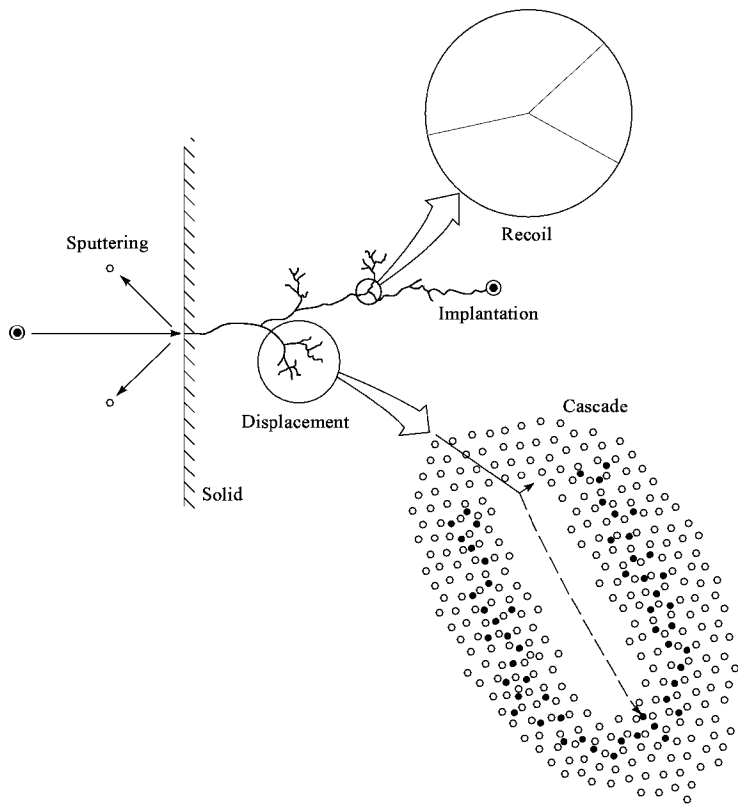


Fig. 1. The ballistic interactions of an energetic ion with a solid. Depicted are sputtering events at the surface, single-ion single-atom recoil events, the development of a collision cascade involving a large number of displaced atoms, and the final position of the incident ion. ○ normal atom; ● interstitial atom; ⊙ incident ion.

The implantation system shown in Figure 2 illustrates a conventional ion implantation system in widespread use within the semiconductor industry. Using different types of available ion sources, a wide variety of beams can be produced with sufficient intensity for implantation processes required for integrated circuit technology. For semiconductors, a representative ion dose is $10^{14} - 10^{15}$ ions cm^{-2} (metallurgical applications generally require doses from $10^{17} - 10^{18}$ ions cm^{-2}). This system produces a unidirectional beam and, in this article, is referred to as a directed beam system. A mass-separating magnet (for mass analysis) is almost mandatory for semiconductor processing in order to eliminate unwanted species that often contaminate the extracted beam. However, for metallurgical processing, mass separation is not important and, as a result, the basic instrumentation can be quite simple.

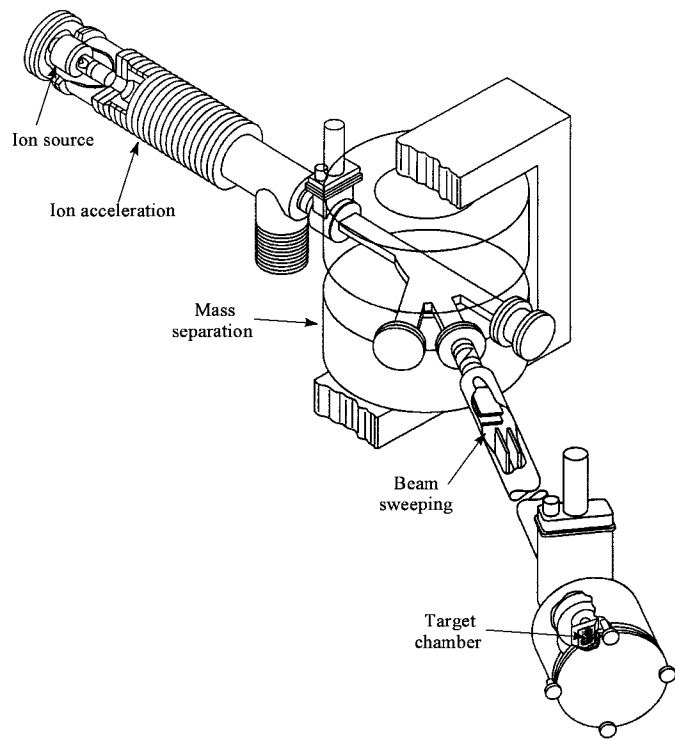


Fig. 2. Schematic drawing of a conventional, directed beam ion implantation system (1).

One ion implantation system which does not use mass analysis and is capable of extremely high ion currents is the broad beam system. Broad beam ion sources (Fig. 3) typically employ grids at the front end of the source to obtain electrostatic acceleration of ions. These sources originated in research programs in the early 1960s as a technology for space propulsion. Since that early work, broad beam systems have been successfully used in the areas of ion implantation, ion beam deposition, and ion beam-assisted deposition where the ions employed are in the low energy range of a few tens of electron-volts (eV) to several thousand eV. Like conventional ion implantation systems, broad beam systems are also referred to as directed beams.

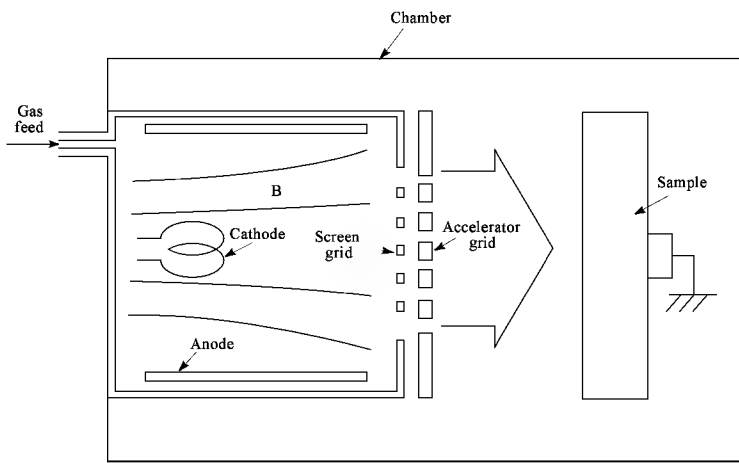


Fig. 3. A typical design for a broad beam, B, ion implantation system.

The plasma source implantation system does not use the extraction and acceleration scheme found in traditional mass-analyzing implanters, but rather the sample to be implanted is placed inside a plasma (Fig. 4). This ion implantation scheme evolved from work on controlled fusion devices. The sample is repetitively pulsed at high negative voltages (around 100 kV) to envelope the surface with a flux of energetic plasma ions. Because the plasma surrounds the sample, and because the ions are accelerated normal to the sample surface, plasma-source implantation occurs over the entire surface, thereby eliminating the need to manipulate nonplanar samples in front of the ion beam. In this article, ion implantation systems that implant all surfaces simultaneously are referred to as omnidirectional systems.

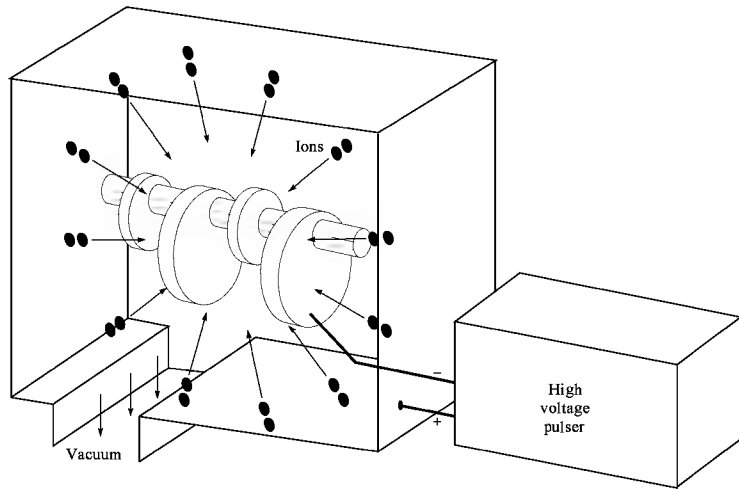


Fig. 4. A schematic of the plasma source ion implantation system, a plasma source chamber linked to a high voltage pulser. The plasma can be created from any gas, often nitrogen. Positive ions strike all surfaces simultaneously.

Ion implantation (outside the traditional semiconductor applications) for the controlled modification of surface sensitive properties has had two principal thrusts: (1) as a metallurgical tool for studying basic mechanisms in areas such as aqueous corrosion, high temperature oxidation, and metallurgical phenomena (eg, impurity trapping); and (2) as a means of beneficially modifying the mechanical or chemical properties of materials. Ion implantation can also modify optical electronic properties of a surface. Optical electrical properties, the traditional industrial application of ion implantation, such as the refractive index, reflectance, conductivity, and magnetic properties can be modified. Chemical properties affected by ion implantation are relevant to the fields of electrochemistry (corrosion), catalysis, and oxidation resistance. The fastest growing research application of ion implantation modifies the mechanical and tribological properties, eg, hardness, modulus, friction, wear resistance, and fatigue resistance, of a material surface.

Some of the advantages of ion implantation in comparison to other surface treatments (such as coatings) are (1) surface properties can be optimized independently of the bulk properties; (2) the process is not limited by thermodynamic constraints, so solid solubility limits can be exceeded by several orders of magnitude, alloy compositions are not limited by diffusion, and metastable compounds can be produced; (3) the process modifies existing surfaces, so there are no interfaces to degrade mechanical properties and original dimensions are retained; (4) low process temperatures avoid thermally related degradation in surface finish and bulk mechanical properties; and (5) the process is highly controllable and reproducible.

Ion implantation processes also have limitations. An intrinsic basic limitation of directed beam ion implantation is that it is a line-of-sight process; it is not feasible to apply it to samples having complicated geometries such as the target shown in Figure 4. Secondly, the range of ions in solids is generally low which leads to shallow penetration and a thin modified layer. Finally, ion implantation as a surface modification tool is generally unfamiliar to most users of other surface modification processes.

These limitations can be addressed in a number of ways. First, plasma source implantation techniques have the ability to treat complicated geometries and are presently being evaluated for commercial applications. Where the estimated cost for beam-line implantation is estimated to be as high as \$0.64 cm² (2) or as low as \$0.01 cm² for coming generation machines (3), industrial-scale plasma source implantation techniques have also been estimated to cost around \$0.01 cm² (4).

The shallow penetration of ion implantation would in itself make it appear useless as a technique for engineering applications; however, there are several situations involving both physical and chemical properties in which the effect of the implanted ion persists to depths far greater than the initial implantation range. The thickness of the modified zone can also be extended by combining ion implantation with a deposition technique or if deposition occurs spontaneously during the ion implantation process. In addition, ion implantation at elevated temperatures, but below temperatures at which degradation of mechanical properties could occur, has been shown to increase the penetration depths substantially (5).

Ion–Solid Interactions

Ion Stopping. Ion–solid interactions are the foundation that underlies the broad application of ion implantation to the modification of materials. The principal features governing the successful exploitation of ion implantation are the range distribution of the energetic ions, the amount and nature of the lattice disorder that is created, and the location of the energetic ions in the crystal lattice. At high dose levels, used to incorporate greater than 5–10 atomic % of implanted species to modify the composition of the target, other phenomena become important: sputtering, ion-induced phase formation, and transformations.

When an energetic ion penetrates a solid, it undergoes a series of collisions with the atoms and electrons in the target. In these collisions the incident particle loses energy at a rate of a few to 100 eV per nanometer, depending on the energy and mass of the ion as well as on the substrate material.

The energy-loss rate of an energetic ion moving through a solid is determined by screened Coulomb interactions with the substrate atoms and electrons. It is customary to distinguish two different mechanisms of energy loss: (1) nuclear collisions, in which energy is transmitted as translatory motion to a target atom as a whole, and (2) electronic collisions, in which the moving particle excites or ejects electrons. For most purposes, this separation into elastic and inelastic collisions is a convenient one and, although not strictly true, it is a good approximation. The energy loss rate dE/dx can be expressed as

$$\frac{dE}{dx} = \left. \frac{dE}{dx} \right|_n + \left. \frac{dE}{dx} \right|_e \tag{1}$$

where the subscripts *n* and *e* denote nuclear and electronic collisions, respectively. Values for dE/dx have been tabulated (6). A schematic energy loss process is shown in Figure 5.

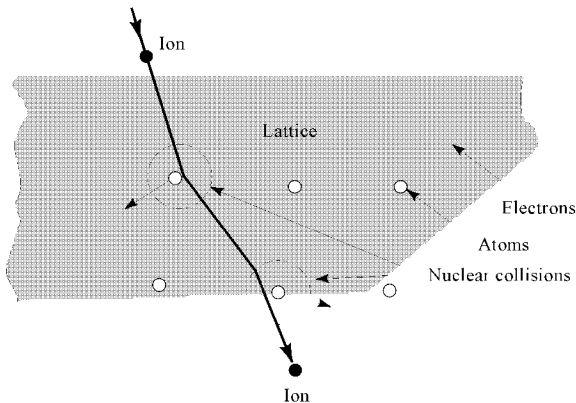


Fig. 5. An ion incident on a crystal lattice is deflected in nuclear collisions with the lattice atoms and also loses energy in collisions with electrons (7).

Nuclear collisions can involve large discrete energy losses and significant angular deflection of the trajectory of the ion. In nuclear stopping, the average energy loss results from elastic collisions with target atoms. This process is responsible for the production of disorder by the displacement of atoms from their lattice position. Electronic collisions occur continuously and involve much smaller energy losses per collision, negligible deflection of the ion trajectory, and negligible lattice disorder. Electronic stopping is an inelastic process and results from energy transferred from the ion to the target electrons. Typical units for the energy loss rate are eV nm or keV μm.

A proper understanding of the mechanisms of energy loss is important not only in controlling the depth profile of implanted dopant atoms, but also in determining the nature of the lattice disorder produced during ion implantation or ion irradiation of the solid. In the process of slowing down in the substrate, the implanted ions undergo violent collisions with some of the lattice atoms, thereby displacing them from lattice sites. Other secondary effects accompanying ion implantation and ion irradiation of solids, such as sputtering of target atoms, also depend strongly on the relative importance of nuclear and electronic stopping. A great deal has been published on the stopping of ions in solids (see *General References*).

Range. Nuclear interactions consist of individual elastic collisions between ion and target atom nuclei, whereas the electronic interactions can be

viewed more as a continuous viscous drag phenomena between the injected ions and the sea of electrons surrounding the target nuclei. For the energy regime normally used in heavy ion implantation (tens to hundreds of keV) the nuclear contribution to the stopping process normally dominates and this is reflected in the particular ion trajectories as the ion comes to rest within the solid. The range R is determined by the rate of energy loss along the path of the ion,

$$R = \int_{E_0}^0 \frac{1}{dE/dx} dE \tag{2}$$

where E_0 is the incident energy of the ion as it penetrates the solid.

Figure 6 shows a two-dimensional schematic view of an individual ion's path in the ion implantation process as it comes to rest in a material. The ion does not travel in a straight path to its final position due to elastic collisions with target atoms. The actual integrated distance traveled by the ion is called the range, R . The ion's net penetration into the material, measured along the vector of the ion's incident trajectory, which is perpendicular to the surface in this example, is called the projected range, R_p .

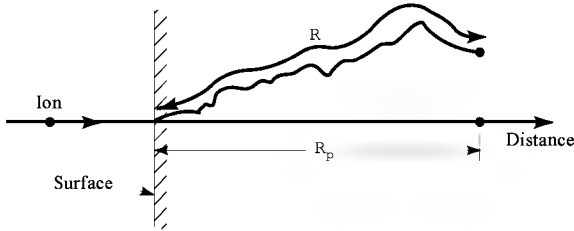


Fig. 6. An incident ion penetrates with a total path length R , which gives a projected range, R_p along the direction parallel to that of the incident ion (1).

Because the stopping of an ion is a random process, the collision sequence and subsequent ion deflection, and the ion's total path length R , vary randomly from ion to ion. As a result, ions with the same energy, incident with the same angle onto the sample surface, and into the same material, do not come to rest in the same place. Hence, all ions of a given type and incident energy do not have the same range. Instead, examination of the range history of many ions shows a statistically broad distribution in the depths to which ions penetrate. The distribution in projected ranges is referred to as the range distribution or range straggling ΔR_p , with the most probable projected range referred to as the average or mean projected range R_p .

In range theory the range distribution is regarded as a transport problem describing the slowing down of energetic ions in matter. Two general methods for obtaining range quantities, one using simulations and the other employing analytical methods, have been developed. The analytic approach used to obtain range quantities is commonly referred to as LSS theory for the pioneering authors (8). Although not precisely accurate, the LSS approach allows calculations of range values with an accuracy of about 20%, which is quite acceptable for most purposes. A more exacted transport calculation is available using the Monte Carlo program TRIM (Transport of Ions in Matter) (9,10). All the methods discussed in this section assume that the target is amorphous and ignore crystal orientation effects.

To discuss the analytical approach to estimating ion ranges, the concept of reduced energy must first be introduced. The reduced energy ϵ is given by equation 3:

$$\epsilon = \frac{E}{Z_1 Z_2 e^2} \frac{a_{TF} M_2}{M_1 + M_2} \tag{3}$$

where E is the particle energy (generally approximated by E_0), $e^2 = 1.44\text{eV}\cdot\text{nm}$, M_1 and Z_1 are the mass and atomic number of the incident particle, M_2 and Z_2 are the mass atomic number of the target atom, and a_{TF} is the Thomas-Fermi screening distance given by equation 4:

$$a_{TF} = \frac{0.88534 a_0}{Z_{\text{eff}}^{1/3}} \tag{4}$$

Z_{eff} is the effective charge number in the interaction of two unlike atoms, and a_0 is the Bohr radius for the hydrogen atom, $0.5292 \times 10^{-8} \text{ cm}$. There exist a number of approximations for Z_{eff} but a simple description based on a mean value is as follows.

$$Z_{\text{eff}} = \left(Z_1^{1/2} + Z_2^{1/2} \right)^2 \tag{5}$$

Simple estimates of range can be obtained using the power law description of nuclear stopping and ignoring electronic stopping. Nuclear stopping is the more important process at low energies, reaching a maximum around $\epsilon = 0.35$, and then falls off with increasing ϵ . Electronic stopping, on the other hand, increases linearly with ion velocity and becomes the dominant process for energies greater than $\epsilon \approx 3$. At intermediate energies, $0.05 < \epsilon < 10$, a rather useful rule-of-thumb for predicting heavy-ion ranges, usually with an accuracy of 30–40%, is equation 6 where ρ is the mass density of the target.

$$R(\text{nm}) = \frac{6E(\text{keV})}{\rho(\text{g cm}^{-3})} \frac{M_2}{Z_2} \frac{M_1 + M_2}{M_1} \frac{(Z_1^{2/3} + Z_2^{2/3})^{1/2}}{Z_1} \tag{6}$$

An approximate measure of the projected range can be found using LSS theory (8) for ions in the energy range where nuclear stopping dominates (eq. 7).

$$R_p \approx \frac{R}{1 + M_2/3M_1} \tag{7}$$

This formula gives values correct to about 15%, but a more exact relation between range and projected range has been calculated using a power law-based LSS theory (11).

The range straggling can be calculated using LSS theory (8) for the condition $\epsilon < 3$ (nuclear stopping dominates) and $M_1 > M_2$ (small angle scattering is favored) (eq. 8): This approximation is good to 20% if the condition $M_1 > M_2$ is satisfied. If $M_1 < M_2$, this approximation is good to only about 40%.

$$\Delta R_p \sim 2.5 R_p \tag{8}$$

The accurate treatment of ion ranges in compound targets requires extensive calculations and is most accurately handled by simulation programs

such as TRIM. However, estimates can be made using two simple techniques. For different atomic species that are sufficiently close in atomic number, ie, Fe–Ti alloys, substitute the mean atomic number and mass into the LSS equations, equations 3 through 5, and proceed as for a monatomic target using equation 6. If the atomic numbers are appreciably different, a first-order estimate may be obtained for an alloy A_xB_y using the expression of equation 9 where $x + y = 1$, $R_p(A)$, $R_p(B)$, N_A , and N_B are the projected ranges and the atomic densities in pure substrates A and B , respectively, and N_{alloy} is the atomic density of the alloy. Equation 9 gives values which are in good agreement with simulations and LSS theory.

$$R_p(A_xB_y) \sim N_{\text{alloy}} \left[\frac{(R_p(A) N_A)(R_p(B) N_B)}{(yR_p(A) N_A) + xR_p(B) N_B)} \right] \tag{9}$$

An estimate of ΔR_p in alloys can be made using the empirical expression (12) of equation 10 where the average alloy reduced energy, ϵ_{av} , is defined by equation 11, where C_i ($i = 1, 2, \dots, n$) is the elemental atomic fraction of the i th element, and ϵ_i is the elemental reduced energy defined in equation 3. Using this formulation, the projected range straggling in compounds can be calculated to within 20%.

$$\frac{\Delta R_p}{R_p} = 0.27 + \frac{0.38}{\epsilon_{\text{av}} + 2.0} \tag{10}$$

$$\epsilon_{\text{av}} = \sum_{i=1}^n C_i \epsilon_i \tag{11}$$

Implanted Species Concentration. The peak atomic density N_p in the ion implantation distribution is estimated using equation 12 where N_p is in units of atoms cm^{-3} , ϕ_i is the ion dose in units of atoms cm^{-2} , and ΔR_p is in units of cm.

$$N_p = \frac{0.4\phi_i}{\Delta R_p} \tag{12}$$

To obtain the peak atomic concentration C_p resulting from this peak number of implanted ions requires knowing N , the atomic density of the substrate. The general relation for the concentration of the implanted species at the peak of the distribution is given by equation 13:

$$C_p = \frac{N_p}{N_p + N} \tag{13}$$

Channeling. All theories examined earlier concerning the ranges of ions and radiation damage of the material were based on the assumption that the stopping medium is disordered, ie, amorphous. Most targets are actually polycrystalline or monocrystalline substances. The main parameters determining the range of an ion are its energy E and atomic number Z_1 , and the atomic number Z_2 of the substrate. In the case of single crystals, the orientation of the substrate and the vibrational amplitude (temperature) of the lattice atoms are also important parameters.

The crystal orientation influence on ion penetration is called channeling or the channeling effect. A schematic comparison of the range distribution under nonchanneling and channeling conditions is shown in Figure 7. When an ion trajectory is aligned along atomic rows, the positive atomic potentials of the line of atoms steer the positively charged ion within the open space, or channels, between the atomic rows. These channeled ions do not make close-impact collisions with the lattice atoms and have a much lower rate of energy loss and hence a greater range than those of nonchanneled ions. The depth distribution of channeled ions is difficult to characterize under routine implantation conditions. The channeling distribution depends on surface preparation, substrate temperature, beam alignment, and disorder introduced during the implantation process itself. An excellent discussion of the channeling effect during ion implantation has been given (13).

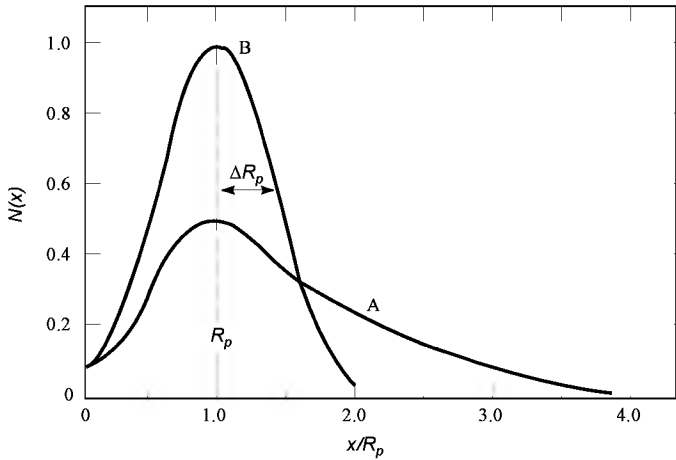


Fig. 7. A shows range distributions for channeled ions implanted along the $\langle 100 \rangle$ axis of Si. B shows the Gaussian distribution for incident ions aligned away from any channeling direction (7).

The channeling effect requires that the incident ions be aligned within a critical angle of the crystal axes or planes. The critical angle depends on the ion energy, ion species, and substrate, but is typically less than 5°. Consequently, the substrate holders for IC processing are often tapered so that the wafers are mounted 7° off normal to minimize channeling effects. However, some ions originally incident at angles greater than the critical angle can be scattered into a channeling direction. It is difficult to avoid channeling effects completely unless the implanted region has been made amorphous by a previous implantation.

Radiation Damage. It has been known for many years that bombardment of a crystal with energetic (keV to MeV) heavy ions produces regions of lattice disorder. An implanted ion entering a solid with an initial kinetic energy of 100 keV comes to rest in the time scale of about 10^{-13} s due to both electronic and nuclear collisions. As an ion slows down and comes to rest in a crystal, it makes a number of collisions with the lattice atoms. In these collisions, sufficient energy may be transferred from the ion to displace an atom from its lattice site. Lattice atoms which are displaced by an incident ion are called primary knock-on atoms (PKA). A PKA can in turn displace other atoms, secondary knock-ons, etc. This process creates a cascade of atomic collisions and is collectively referred to as the collision, or displacement, cascade. The disorder can be directly observed by techniques sensitive to lattice structure, such as electron-transmission microscopy, MeV-particle channeling, and electron diffraction.

Collision cascades (see Fig. 1) lead to a distribution of vacancies, interstitial atoms, and other types of lattice disorder in the region around the ion

track. As the number of ions incident on the crystal increases, the individual disordered regions begin to overlap. At some point a heavily damaged layer is formed. The total amount of disorder and the distribution in depth depend on ion species, temperature, energy, total dose, and channeling effects. The average number of displaced atoms in a cascade produced by a PKA of energy E is denoted by $\langle N_d(E) \rangle$, also known as the displacement damage function. By correctly accounting for electronic stopping and using a realistic interatomic potential to describe atomic interactions, the damage function is given by (14–16) equation 14 where E_d is the displacement energy (typically about 25 eV) and $\nu(E)$ is the amount of PKA energy not lost to electronic excitation, commonly referred to as the damage energy. The damage energy in reduced notation can be approximated as shown in equation 15 for $\epsilon < 1$ (see eq. 3) and $Z_1 > 5$.

$$\langle N_d(E) \rangle = \frac{0.8\nu(E)}{2E_d}$$

(14)

$$\nu(E) \sim 0.8E$$

(15)

The simplest calculation of radiation damage involves only monatomic materials and has been described by many authors (17–20). For polyatomic materials, a calculation procedure for estimating damage energy from ion implantation has been outlined (8). The extension of this formalism (8) to direct calculations of damage energies in polyatomic materials has been addressed by several authors (11,21–24). A commonly used measure of irradiation damage is displacements per atom (dpa). A unit of 1 dpa means that on the average, every atom in the irradiated volume has been displaced once from its equilibrium lattice site. The approximated dpa in the implanted region is given by equation 16 where θ (ions cm^{-2}) is the dose, and $\langle N_d(E) \rangle$ is the damage function given by equation 14.

$$\text{dpa} \sim \frac{\theta \langle N_d(E) \rangle}{N R_p}$$

(16)

Radiation Enhanced Diffusion. Ion irradiation is quite efficient in forming vacancy–interstitial pairs. The atomic displacements resulting from energetic recoiling atoms can be highly concentrated into small localized regions containing a large concentration of defects well in excess of the equilibrium value. If the defects are produced at temperatures where they are mobile, and can in part anneal out, the balance between the rate of formation vs the rate of annihilation leads to a steady-state concentration of defects. Since the atomic diffusivity is proportional to the defect concentration, an excess concentration of defects leads to an enhancement in the diffusional process (25–27). The defects generated in ion–solid interactions influence the kinetic processes that occur both inside and outside the cascade volume. At times long after the cascade lifetime ($t > 10^{-11}$ s), the remaining vacancy–interstitial pairs can contribute to atomic diffusion processes. This process, commonly called radiation enhanced diffusion (RED), can be described by rate equations and an analytical approach (27). Within the cascade itself, under conditions of high defect densities, local energy depositions exceed 1 eV/atom and local kinetic processes can be described on the basis of a liquid-like diffusion formalism (28,29).

Sputtering. The erosion of a surface by energetic particle bombardment is called sputtering. In this process surface atoms are removed by collisions between the incoming particles and the atoms in the near surface layers of a solid (see Fig. 1). Sputtering sets the limit of the maximum concentration of atoms that can be implanted and retained in a target material. The yield of sputtered atoms Y , the number of sputtered atoms per incident ion, is typically between 0.5 to 20 depending on ion species, ion energy, target material, and angle of incidence of the ion onto the target material. An extensive list of sputtering yields has been given (30).

The generally accepted theory (31) that explains most sputtering phenomena in elemental materials is based on the collision cascade picture. The energy of the incident ion is shared among those atoms within the collision cascade volume and then dissipated. Only those collisions that take place near the surface of the material are directly effective in knocking atoms out of the material. The majority of sputtered atoms emerge only from the first few atomic layers. Therefore, the more collisions taking place in the near-surface region, the higher the sputtering yield. In addition, the sputtering yield is proportional to the nuclear stopping power of the incident ion in the near-surface region. In other words, the higher the incident ion energy, the lower the nuclear stopping power is near the surface and the lower the sputtering yield.

The sputtering yield is proportional to the number of displaced atoms. In the linear cascade regime that is applicable for medium mass ions (such as argon), the number of displaced atoms, $F_D(E_o)$, is proportional to the energy deposited per unit depth as a result of nuclear energy loss. The sputtering yield Y for particles incident normal to the surface can be expressed as follows (31).

$$Y = \Lambda F_D(E_o)$$

(17)

The material factor Λ contains the material parameters and is a description of the number of recoil atoms that can escape from the solid. In one description (31) (eq. 18), N is the atomic density of target atoms and U_o is the surface binding energy.

$$\Lambda \simeq 4.2 \frac{N U_o}{E_o}$$

(18)

U_o can be estimated from the cohesive energy and has typical values between 2 and 4 eV. Values of the cohesive energy have been given (32). $F_D(E_o)$ can be expressed as in equation 19 for $\epsilon \ll 1$ and $Z_1 > 5$. In this equation α is a correction factor, which is a function of M^2/M^1 and has values between 0.2 and 0.7.

$$F_D(E_o) \simeq \frac{\alpha 0.8 E_o}{R_p}$$

(19)

When the bombarding ion is incident at glancing angles, the sputtering yield differs from the normal incidence yield. In general, it is observed that the sputtering yield for an incidence angle θ , $Y(\theta)$, is related to the normal incidence sputtering yield $Y(0)$ according to equation 20 where θ is measured from the surface normal and the exponent f_s is approximately 2 (33).

$$\frac{Y(\theta)}{Y(0)} = (\cos \theta)^{-f_s}$$

(20)

Sputtering effects also give a strong angular dependence to the retained dose of ion implanted profiles. Experimentally, this dependence has been seen in a number of cases including measurements of retained doses for high fluence implantations (ie, $> 1 - 3 \times 10^{17}$ ions cm^{-2}) in metals. Figure 8 shows such a measurement for high dose ion implantation into metals. The extreme drop-off of the retained dose with angle can be seen. This exemplifies the necessity of implanting at near normal angles of incidence to maximize retained dose.

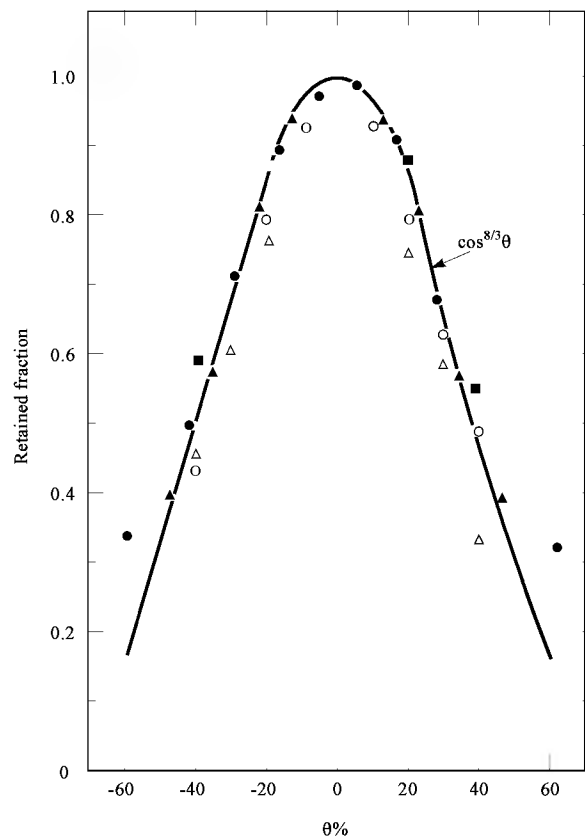


Fig. 8. Normalized retained dose at saturation from simplified theory and from measurements taken on cylindrical steel cylinders (34). Symbols, targets, and $\phi_o(10^{17} \text{ cm}^2)$ are as follows: ■, Ar, 52100, 3; ●, Ti, 304, 10; ▲, Cr, M50, 2; ○, Ti, 52100, 3; and Δ, Ta, M50, 1.

Since ion implantation is also a bombardment of energetic ions, there is always some sputtering, especially when heavy ions and high doses are used. Sputtering makes the sample surface recede as shown schematically in Figure 9a; therefore it affects the implantation profile (Fig. 9b) and also removes atoms that have been implanted. This eventually leads to a steady-state condition (eq. 21) in which there is no further increase in the amount of implanted species retained in the material, (36), where N_A and N_B are the concentrations (per unit volume) of A and B atoms, Y is the total sputtering yield, and r is the ratio of the probability for a B atom near the surface to be sputtered to that of an A atom to be sputtered. This is the steady-state surface composition of the implanted ion, atom A , and the monatomic target, atom B . For direct ion implantation into a target material, the maximum concentration of implanted species is inversely proportional to the sputter yield. This maximum concentration is obtained after the sputtering of a thickness comparable to the ion range, R_p (or more exactly, $R_p + \Delta R_p$). However, more careful consideration should be given if there is preferential sputtering between atoms of the host material and those of the implanted species. This implies that one has to sputter an amount of material, equal to r times the thickness R_p , in order to reach the steady state. Consequently for ion-target combinations with high sputter yields, the maximum concentration may be only a few atom %.

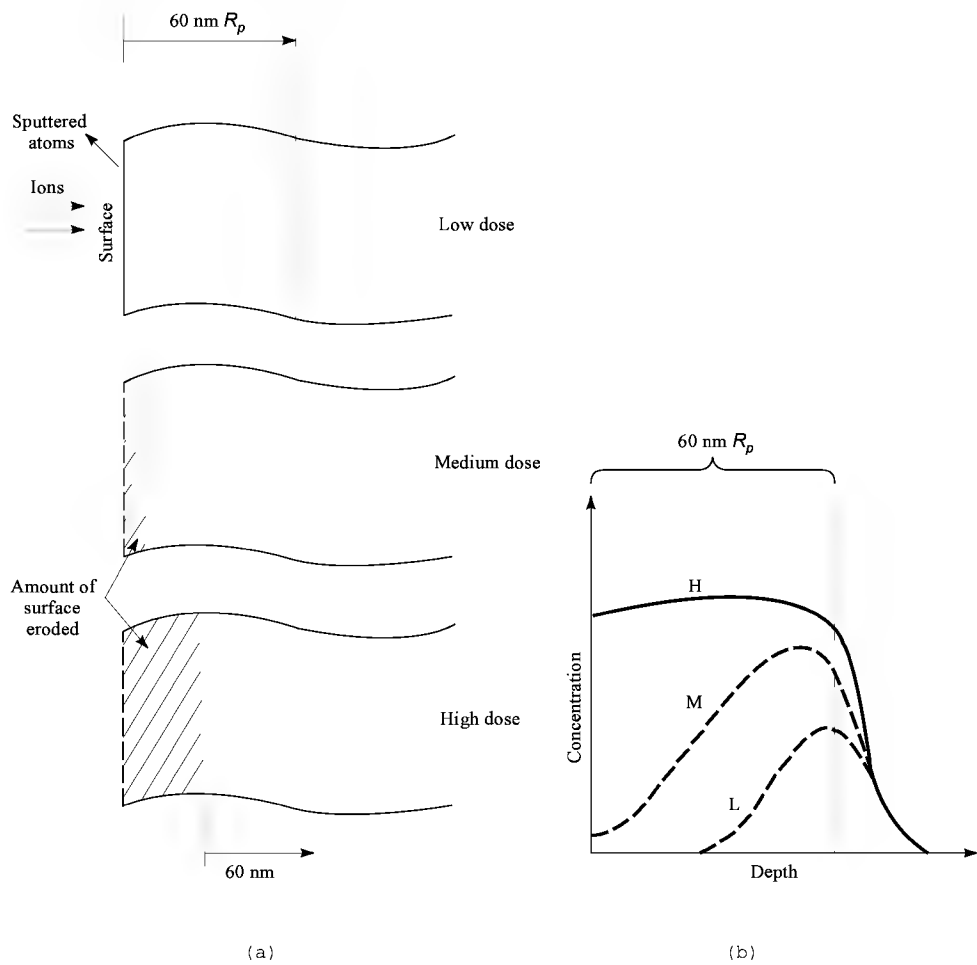


Fig. 9. Schematic view of the development of the concentration profile of ions implanted from low (L), medium (M), and high (H) doses. The projected range in this sample is 60 nm. (a) Surface erosion; (b) implantation profile (35).

$$N_A/N_B = r(Y-1)^{-1}$$

(21)

The main features of sputtering remain the same for composite materials such as binary alloys. However, there are additional complications because there are two kinds of atoms in the material. The two species may not be sputtered at an equal rate because of differences in energy sharing (in the collision cascade), ejection probabilities, or binding energies. Preferential sputtering of one species over the other has been observed in many alloys and compounds. The change in surface composition of sputtered multicomponent targets is well documented (37) and has been theoretically described (38).

Since the majority of sputtered atoms have relatively low energies and emerge from the first few atomic layers near the surface, the probability of sputtering is very sensitive to surface conditions. Surface conditions are influenced by several factors, such as residual gas in the vacuum, target material, and current density of the incident ion beam. For example, it is well known that ion implantation in a bad vacuum can cause the formation of a carbon layer on the sample surface. Formation of thin oxide layers are often encountered in the sputtering of easily oxidized materials. Good vacuum and high ion-beam current (high sputtering rate) are often desirable to minimize surface oxidation.

A thin layer of surface contaminants or oxide can effectively protect the material from being sputtered, and therefore can strongly affect the parameters Y and r , which in turn determine the state of the implanted materials. Since both carbon and oxide layers can greatly reduce the sputtering yield of the material, they can significantly increase the maximum implanted concentration (39–42). It might appear desirable to have surface oxide and carbon layers intentionally to enhance the implant concentration. However, because of atomic mixing, the surface oxygen and carbon can be mixed into the implanted layer after the prolonged implantation. Significant side effects, sometimes undesirable, can be caused by these impurities.

Sputtering can also give rise to surface roughness, which can possibly affect a high dose implantation. Surface roughness has been found to be related to crystallographic orientation, impurities in the material, ion species, and angles used for sputtering. An extremely rough surface can also reduce the sputtering yield.

Simulations. In addition to analytical approaches to describe ion–solid interactions two different types of computer simulations are used: Monte Carlo (MC) and molecular dynamics (MD). The Monte Carlo method relies on a binary collision model and molecular dynamics solves the many-body problem of Newtonian mechanics for many interacting particles. As the name Monte Carlo suggests, the results require averaging over many simulated particle trajectories. A review of the computer simulation of ion–solid interactions has been provided (43).

The Monte Carlo methods, applied to ion–solid interactions, have a number of distinct advantages over analytical calculations based on transport theory. The MC approach allows for a more rigorous treatment of elastic scattering and of the determination of angular and energy distributions. A number of MC codes have been developed over the years (43). The various MC programs differ primarily in their basic treatment of elastic scattering. The program TRIM (transport of ions in matter) (9,10) is the most commonly cited for range and damage distributions in amorphous materials. TRIM can also simulate sputtering processes. The program provides high computer efficiency, and the agreement between TRIM and experimental data is excellent. The influence of the crystal lattice on the range and damage distributions has been accounted for in the MC program called Marlowe (44–46).

To examine the solid as it approaches equilibrium (atom energies of 0.025 eV) requires molecular dynamic simulations. Molecular dynamic (MD) simulations follow the spatial and temporal evolution of atoms in a cascade as the atoms regain thermal equilibrium in about 10 ps. By use of MD, one can follow the physical and chemical effects that influence the final cascade state. Molecular dynamics have been used to study a variety of cascade phenomena. These include defect evolution, recombination dynamics, liquid-like core effects, and final defect states. MD programs have also been used to model sputtering processes.

The primary requirement for carrying out MD simulations is a suitable interatomic potential for the description of forces acting between atoms in the cascade. A general discussion on MD can be found (47) and detailed summaries of the use of MD in ion–solid interactions are also available (43,48). An extended discussion on embedded-atom potentials necessary for MD calculations can also be found (49).

Other Processes Utilizing Ion Beams. Materials under ion irradiation undergo significant atomic rearrangement. The most obvious example of this phenomenon is the atomic intermixing and alloying that can occur at the interface separating two different materials during ion irradiation. This process is known as ion beam mixing (IBM). The advantage of IBM is that arbitrary concentrations of the materials are readily attainable, and the composition of the surface can be controlled independent of the bulk materials. IBM coatings are generally less than 100 nm thick, approximately equivalent to the range of a typical implanted ion. An early observation of the ion mixing phenomenon was made following the irradiation of a silicon substrate coated with a thin palladium film (50). Further information about ion beam mixing can be found (51–55).

A related process uses ions to bombard material as it is being deposited onto a substrate. This process is called ion beam assisted deposition (IBAD) or ion assisted deposition (IAD). Interest started turning to this hybrid technique in the early 1980s and continues to the present (1994). IBAD films can overcome the thickness limitations of ion implantation and IBM and still maintain desirable adhesion. There are several reviews on this (56–63). Many of the original researchers exploring IBAD processing were primarily interested in better understanding the processes in plasma-based deposition techniques and deposition of semiconductor-related films. Other application areas, such as corrosion and tribology, are now being given more attention and are expected to accelerate in their development.

A number of technologically important coatings have been deposited using the IBAD process. These include diamond-like carbon (DLC) (64–67), boron nitride (68–70), titanium nitride (71), group IVB nitrides (72), dielectric coatings (73–75), optical coatings (74,76–77), reflective coatings (78), thermochromic coatings (79), magnetic thin films (80), tribological hard coatings such as titanium nitride (81–83), titanium carbide (84), chromium nitride (84), and cubic boron nitride (85), solid lubricant coatings (86–87), and aqueous corrosion-resistant coatings (88–90).

Ion Implantation Applications

Tribology. The term tribology (the science and technology of interacting surfaces in relative motion) encompasses an old, important, and often complicated set of phenomena relating to friction and wear. In general, the wear properties of materials are more important than their friction properties. Implantation has thus been mostly studied for improving wear resistance (91–100). The application of ion implantation to these areas has not only yielded surfaces with improved properties but has been important for the study of basic mechanisms. The problem areas impacted by implantation can be grouped into three main categories: (1) cutting and slitting operations, (2) corrosive applications and adhesive wear, and (3) extrusion operations and applications where large surface forces occur. Industrial applications have been reviewed (101,102). Two excellent handbooks on tribology which include ion beam processing are available (103). A review of several implantation topics being investigated in Japan and elsewhere has appeared (104).

Ion implantation of metals is becoming more routine on a commercial scale, mainly with nitrogen implantation as an antiwear treatment of high value critical components. The primary use is for the antiwear treatment of surgical prostheses such as hips and knees. In use, these implanted components are in articulating contact against a mating ultrahigh molecular weight polyethylene cup and wear of either component is a prime concern for the longevity of the (surgical) prosthesis. In 1994 approximately 100,000 knee, hip, and other joint prostheses were being treated each year in the United States (105). Ion implantation for such medical devices is attractive since there is no concern regarding delamination as for sharp interfaces, and nitrogen is considered benign in the human body.

Ion implantation appears to be an attractive technique for treating industrial components by (1) the stabilization of microstructure, preventing a change in wear mode, (2) the stimulation of transformations to a wear-resistant mode, or (3) the creation of chemical passivity to prevent a corrosive wear mode. Components benefiting from nitrogen ion implantation for improving wear resistance of tool steel alloys include plastic injection molding tools, metal rolls, piercing tools, forming tools, and other components used in mild-wear applications. Successful utilization of nitrogen implantation requires

relatively low tool surface operating temperatures since the nitrogen defect structures attributed to improvement of wear resistance are not stable at high temperatures. The implantation doses typically used for these applications range from 2 to 6×10^{17} ions cm^{-2} at energies of 50–100 keV.

The gap between laboratory wear testing and industrial application trials is extremely difficult to bridge, since there is often little or no control over testing in the industrial environment. Despite these limitations, several examples of industrial successes involving ion implanted tools have been reported and blind tests of nitrogen-implanted machine tools have been performed, including tool taps, dies, punches, and TiN coated WC cutting inserts (106). The implanted tools showed lifetime improvements ranging from 1.5× to 4× and no unimplanted tool demonstrated better performance than an implanted tool. Improvements were also observed for implanted tool dies and punches.

Fatigue. Fatigue represents a singularly dangerous mode of material failure, in that no obvious prior warning is given of impending fracture. Generally, such failure occurs upon the cyclic loading at some stress below the static fracture stress. High loading amplitudes give rise to short lifetimes (low cycle fatigue) whereas relatively low loads yield longer lifetimes (high cycle fatigue).

Ion implantation has been employed to improve high cycle fatigue in copper (107), steel (108,109), nickel (110), and titanium alloys (111–116). The improvements for low cycle fatigue are smaller. In both cases, ion implantation changes near surface slip, promoting reversibility and increasing the homogenization or suppressing surface slip. Strengthening mechanisms involved include solid solution hardening, precipitation hardening, and the production of compressive stress. The failure mode of implanted surfaces is seen to shift from slip band cracking to grain boundary cracking. This topic has been reviewed (117).

High temperature fatigue and fretting fatigue behavior has also been improved by implantation (113,114). This has been achieved by using species that inhibit oxidation or harden the surface. It is generally accepted that fretting behavior is closely connected to oxidation resistance, perhaps due to third party effects of oxidation products. Oxidation resistance alone has also been improved by ion implantation (118–120).

Aqueous Corrosion. Several studies have demonstrated that ion implantation may be used to modify either the local or generalized aqueous corrosion behavior of metals and alloys (119,121). In these early studies metallic systems have been doped with suitable elements in order to systematically modify the nature and rate of the anodic and or cathodic half-cell reactions which control the rate of corrosion.

The following mechanisms in corrosion behavior have been affected by implantation and have been reviewed (119): (1) expansion of the passive range of potential, (2) enhancement of resistance to localized breakdown of passive film, (3) formation of amorphous surface alloy to eliminate grain boundaries and stabilize an amorphous passive film, (4) shift open circuit (corrosion) potential into passive range of potential, (5) reduce eliminate attack at second-phase particles, and (6) inhibit cathodic kinetics.

Corrosion studies involving implantation also fall into two categories: (1) studies of novel surface alloys (119,121–123) and (2) attempts to improve the corrosion resistance of some commonly used engineering alloys (124,125). The nature of the microstructure of a surface alloy can have a significant influence on corrosion behavior. It is well known that multiphase alloys tend to be susceptible to localized galvanic corrosion between phases of different chemical reactivity. Thus it is always desirable to produce single-phase alloys to avoid such effects. Chemical homogeneity in single-phase alloys is also desirable. Ion implantation may be used to form single-phase solid solutions often far in excess of the equilibrium composition. From the corrosion scientist's viewpoint this is one significant advantage of the use of ion implantation as a surface alloying technique.

In addition to the conventional approach to designing corrosion-resistant alloys, ion implantation offers a method for the formation of amorphous surface alloys (119,126). It is known that amorphous alloys formed by rapid quenching often exhibit superior corrosion resistance provided that the alloy has a sufficient concentration of a strong passivator such as chromium. One advantage of such alloys is that the absence of grain boundaries allows for the formation of a continuous passive film which is not disrupted at the grain boundary region. The majority of highly corrosion-resistant alloys studied so far conform to a composition of approximately 80 atomic % transition elements and 20 atomic % metalloids.

The most researched application for corrosion resistance by ion implantation has been for high cost, high precision aerospace bearings (127). This thrust has been due to the promise of no delamination problems and the possibility of selectively optimizing the surface of the tough, hardened martensitic bearing alloys whose bulk is optimized for high load carrying capabilities and rolling contact fatigue. There is a widespread bearing corrosion problem in military aircraft propulsion systems which is typified by localized pitting along the contact region between the rollers and races. The corrosion pits may act as initiation sites for fatigue spalling which can lead to catastrophic engine failure. Another serious problem is that replacement bearings can have short shelf-lives, again due to corrosion. Ion implantation was applied to AISI M50 and AISI 52100 bearing steels (125,128). They found that in addition to maintaining dimensional stability, mechanical integrity in the form of rolling contact fatigue resistance was not altered by the implantation process.

Catalysis. Ion implantation and sputtering in general are useful methods for preparing catalysts on metal and insulator substrates. This has been demonstrated for reactions at gas–solid and liquid–solid interfaces. Ion implantation should be considered in cases where good adhesion of the active metal to the substrate is needed or production of novel materials with catalytic properties different from either the substrate or the pure active metal is wanted (129–131). Ion beam mixing of deposited films also promises interesting prospects for the preparation of catalysts (132).

Reactions studied at solid–liquid interfaces have been concerned with electrocatalysts (electrodes) mainly in systems important for the development of fuel cells or water electrolysis. In a model study implanting platinum and other metals into iron electrodes, a three orders of magnitude increase of the hydrogen evolution rate in acidic solutions, compared to unimplanted iron and more than two orders of magnitude in comparison to smooth platinum, was demonstrated (131,133). In contrast to the nearly inactive implanted case, the activity of the catalyst prepared by sputtering was higher than for the smooth platinum metal. An interesting additional finding was the further increase of the activity, by nearly one order of magnitude, after intermixing the sputtered platinum layer with the substrate by means of an argon ion beam.

Ion implantation has also been used for the creation of novel catalytically active materials. Ruthenium oxide is used as an electrode for chlorine production because of its superior corrosion resistance. Platinum was implanted in ruthenium oxide and the performance of the catalyst tested with respect to the oxidation of formic acid and methanol (fuel cell reactions) (131). The implantation of platinum produced of which a catalytically active electrode, the performance of which is superior to both pure and smooth platinum. It also has good long-term stability. The most interesting finding, however, is the complete inactivity of the electrode for the methanol oxidation.

There are, however, continuing difficulties for catalytic applications of ion implantation. One is possible corrosion of the substrate of the implanted or sputtered active layer; this is the main factor in the long-term stability of the catalyst. Ion implanted metals may be buried below the surface layer of the substrate and hence show no activity. Preparation of catalysts with high surface areas present problems for ion beam techniques. Although it is apparent that ion implantation is not suitable for the production of catalysts in a porous form, the results indicate its strong potential for the production and study of catalytic surfaces that cannot be fabricated by more conventional methods.

Ceramics. The primary approaches to investigating ion implantation in ceramics has been to examine (1) fracture toughness of these materials (134), (2) microhardness (135), and (3) microstructure (136). The behavior of ceramics is in general more complicated than metals and has produced results very dependent on the ceramic and ions examined, and on post-bombardment thermal annealing. Several reviews have been published on this subject (134,136–139). It should also be noted that the implantation of ceramic TiN coatings has a pronounced effect on extending their lifetime when used for tool coatings and has become an important industrial application following large-scale industrial tests (106).

Polymers. Ion implantation of polymers has resulted in substantial increases of electrical conductivity (140), surface hardness (141), and surface texturing (142). A four to five order of magnitude increase in the conductivity of polymers after implantation with 2 MeV Ar ions at dose levels ranging from 10^{15} – 10^{17} ions cm^{-2} has been observed (140). The hardness of polycarbonate was increased to that of steel (141) when using 1 MeV Ar at dose levels between 10^{15} – 10^{16} ions cm^{-2} . Conductivity, oxidation, and chemical resistance were also improved. Improvements in the adhesion of metallizations to Kapton and Teflon after implantation with argon has been noted (142).

Metastable Compound Formation. Metastable alloys are ordinarily formed under conditions involving a rapid thermal quench, such as in splat cooling where effective quenching rates of 10^6 – 10^7 K s can be obtained. Typical techniques capable of obtaining rapid thermal quenching include liquid splat cooling (10^5 – 10^8 K s), vapor condensation (10^{12} K s), and ion beam induced cascade quenching. The stopping of ions can be viewed conceptually as similar on an atomistic scale. Individual atoms within a collision cascade, displaced in the wake of an ion, can have high kinetic energies (several eV) corresponding to extremely high equivalent temperatures, and can dissipate this energy in time periods of 10^{-11} to 10^{-12} s yielding very high local quenching rates (10^{11} – 10^{13} K s) (27,143,144).

Diamond-Like Carbon. The ultrafast quenching times associated with the decay of ion collision cascades have been utilized for producing various classes of metastable compounds, including diamond-like carbon (DLC) coatings (65–70,145–149). These ion-induced coatings were first documented in 1971 (145) when sputtering deposited energetic carbon ions onto substrates resulted in hard coatings being formed. The resulting DLC films were highly adherent and insulating and demonstrated high chemical resistance to acids, bases, and solvents. Studies revealed evidence of microcrystallite diamonds in an amorphous structure with varying amounts of *sp*² (graphitic) and *sp*³ (diamond) bonding (65). These films also contained various amounts of hydrogen which dictated their relative optical and mechanical behavior. They were very hard, inert, and generally in high compressive stress, a factor limiting their use in thick film or free standing use. DLC technology has been reviewed (149–152).

Ion Implantation Technology

Although there has been theoretical and experimental interest in the effects of ion bombardment on materials since about 1960 (153), the growth in ion implantation technology and applications since then is due almost solely to the semiconductor (integrated circuit) industry. The advantages of ion implantation for semiconductor doping were first pointed out in 1955 (154), but these advantages were not widely accepted until about 1970.

Outside of the semiconductor industry, ion implantation has been explored for beneficially modifying surface sensitive properties since the early 1970s. A large share of the early work in this field was performed at Harwell Laboratories, within the U.K. Atomic Energy establishment, with emphasis on tribological properties as modified by nitrogen implantation and oxidation resistance. Several other laboratories worldwide became engaged in ion implantation research subsequently and the range of topics explored expanded to cover other topics and materials, ie, ceramics and polymers.

Directed Beam Ion Implanters. The earliest ion implanters evolved from the isotope separators of the 1940s and later. Ion implanters are frequently classed according to their ion current capabilities, ranging from low currents (microamperes) to high currents (one to several milliamperes). The specific design criteria have been mainly driven by the particular dose and depth profile requirements for semiconductor device fabrication. The history of ion implanter evolution and development is in itself an interesting study of technology transfer. It is covered comprehensively in a series of conference proceedings (see *General References*) that parallel the developments in silicon device technology that has demonstrated such explosive growth since the early 1970s (155). Although this area of accelerator application is not the focus of this article, many ion implanters used for general materials science studies and R&D are either converted semiconductor ion implanters or are based on their design. Therefore, the basic design and features of these systems will be briefly discussed.

The typical (directed beam) ion implanter normally consists of (1) an ion source, (2) an electrostatic accelerator, (3) a mass analysis section (normally a magnet), (4) a means of scanning the ion beam over the target, and (5) a target chamber to hold the substrate to be implanted along with a means of monitoring the ion dose delivered to the substrate. The footprint of a typical medium current (500–1000 μ A current level) ion implanter is shown schematically in Figure 2. Broad beam implanters are of similar construction, but consist only of an ion source and a target chamber (Fig. 3).

The performance and reliability of an implanter's ion source largely determines the system's commercial viability. As such, the majority of commercial (semiconductor) ion implantation ion sources have been based on isotope separator source design. More recent sources are based on filamentless designs promising longer lifetimes and higher currents (156). These include radio frequency (RF) and electron cyclotron resonance (ECR) type ion sources. In any case, the objective is to obtain a low maintenance ion source that is (relatively) long lived, stable, and delivers an intense beam of positive ions of the desired species with a minimum of unwanted species. Most plasma-based heavy-ion sources produce ions utilizing a gaseous discharge at low pressure containing the elemental species of interest (157). The positive ions are extracted from the plasma by means of an electric field (between the extraction electrode and source) and are subsequently focused and directed through a mass analysis magnet to obtain adequate ion purity that is crucial to semiconductor processing but generally less so for materials processing. The advantages and limitations of the predominant sources used for commercial directed beam systems are outlined in Table 1.

Table 1. Comparison of Ion Source Types Used in Directed Beam Ion Implantation

Ion source type	Advantages	Disadvantages	References
arc discharge	high beam currents, stable operation, high efficiency, easily modified	source life is limited by filaments	158–160
metal vapor vacuum arc (MEVVA)	no gas load, high charge states, relatively pure beams (no mass analysis)	unknown charge state distribution, minor impurities expected in nonanalyzed mode	161–162
electron cyclotron resonance (ECR)	high beam currents, low operating pressures	extraction geometries	163

Mass analysis of accelerated ions is imperative for semiconductor applications of ion implantation because of the extreme sensitivity to impurities and is also normally used for materials modification studies using modified semiconductor equipment. Exceptions are instances, such as above, where the ion source is capable of producing relatively pure beams, ie, the MEVVA source. In other cases, such as certain tribological applications involving nitrogen ion implantation, the presence of low levels of impurities is not regarded as a problem. However, in almost all cases when required, magnetic separation is employed for mass analysis in each of the configurations described (164).

Directed beam implanters normally require a drift space for the ion beam if electrostatic scanning is used to sweep the beam across the substrates to ensure uniformity. Beams from broad beam sources have relatively large diameters and are generally not scanned. Alternative methods employed in semiconductor machines include movement of the (wafer) substrates in front of a stationary beam which is required for very high beam currents. This is due to the fact that positively charged ions mutually repel one another during their travel down the beam line, ie, the so called space-charge blow-up. This effect is most prominent at low energies (lower velocities) and for heavy masses and becomes troublesome when employing electrostatic steering or focusing. This drift space also allows the neutralization of ions during their flight which must be adjusted for by appropriate offset of the substrate target chamber.

Another important consideration for providing uniform implantation involves the geometry of the ion beam with respect to the target surface. Too high an angle from normal incidence leads to excessive sputtering and low retained dose. These issues and others pertinent to practical aspects of implantation treatment have been discussed (35,165).

The ultimate goal of using ions for surface modification is dependent on the ability to measure the delivered doses (ions cm^{-2}) and predict the resultant impurity depth profiles. As already discussed, the latter are dictated by well-documented range–energy relationships as well as by sputtering effects which determine the equilibrium retained dose. An accurate determination of the ion beam current is accomplished by utilizing a Faraday cup device capable of collecting and measuring the ion charge with adequate precautions taken to account for secondary electrons produced whenever an ion strikes a surface. Accurate current density measurements require the elimination of effects of secondary electrons present in the ion beam itself as well as those produced as the result of the ion bombardment of surfaces. Since each ion can typically produce more than one secondary electron in a single collision, this correction is often a significant challenge for obtaining high accuracy dose determinations. In a brief review of the literature on secondary electron emission for directed beam experiments, the implications for omnidirectional ion implantation are also discussed (166).

Omnidirectional Ion Implanters. The plasma source ion implantation (PSII) (167,168) process is illustrated in Figure 4. A typical PSII system consists of an ion source, a target chamber, and circuitry and power supplies for pulse biasing the target. A negative high voltage pulse is applied to a target immersed in a plasma. Plasma ions are accelerated by the electrical potential and are implanted into the surface of the target. PSII offers several improvements over conventional techniques. First, particle accelerators are eliminated. Second, PSII is a nonlinear-of-sight process enabling conformal implantation, ie, ions are accelerated from all directions simultaneously into all exposed surfaces of the target. Consequently, target manipulation fixtures, beam rastering, and masking of selected areas of the target are unnecessary. Implant times are short when compared to directed beam techniques since high

current, pulsed-power supplies compatible with this process can provide two orders of magnitude higher average currents than conventional accelerators. Since large areas can be implanted concurrently, ion current densities to the target can be kept low to avoid overheating problems sometimes encountered in directed beam implants (2).

Development of adequate ion sources are required for large-scale implementation of PSII. Gaseous discharges with either thermionic, radio frequency, or microwave ionization sources have been successfully used to produce ions. The production of large-scale, uniform, mass analyzed plasmas is usually technologically and economically prohibitive. Consequently, PSII often produces a broadened implant profile due to the varied stopping ranges of the assorted ion masses that are usually present. For example, when using typical nitrogen plasmas, the atomic ions (N⁺) implant deeper than the molecular ions (N₂⁺). Nongaseous sources such as vacuum arc discharges (169) appear promising, although large, high current steady-state sources must still be developed for a practical PSII system.

Circuitry and power supplies capable of handling the currents and voltages present in large-scale PSII systems are already available. The high voltage pulse networks either generate the voltage directly or consist of a pulse forming network and a step-up transformer. The high voltage technology available for PSII processes has been discussed (170).

Much of the PSII research has been performed to show proof-of-principle on relatively small devices with acceleration voltage magnitudes up to 150 kV. Large-scale demonstration facilities are also in use. The largest facility to date (1994) is operational at the Los Alamos National Laboratory (171), and consists of a 1.6-m diameter, 5-m long plasma processing chamber, powered by a 125-kV, 60-A pulsed current (2.4-A average current) power supply. While much of the research has focused on metallurgical applications, semiconductor processing with PSII, such as high throughput, low voltage, shallow trench implants for submicrometer electronics circuitry, is also being investigated (172).

Target Temperature. Since essentially all of the energy of the stopping ions is ultimately converted into heating of the substrate, it is often essential to provide adequate heat removal during implantation in order to limit the temperature excursions of a particular material, such as a tool steel or temperature sensitive polymeric material. Although ion implantation is capable of being a low temperature process, this depends critically on its thermal conductance to a heat sink. The temperature rise of a material receiving beam power in vacuum depends on its thermal conductance and radiative emission losses. Radiation losses scale with the fourth power of the absolute temperature (K), so this contribution only becomes significant for temperatures above 300–400°C. Therefore, one normally should enhance conduction cooling to limit beam-induced heating during implantation.

Several methods have been employed to optimize conduction cooling, including using malleable foils between the target and the cooled backplate as well as using thermally conducting adhesives. The use of thin metal films to enhance the heat transfer between the target and a thermally controlled heat sink is common. There is a fairly sharp maximum in thermal conductivity, at about 100 μm thickness, due to a competition between malleability (hardness) and conductivity. Several workers have also successfully used low vapor pressure organic substances, such as vacuum greases as well as low melting point eutectics to thermally couple substrates to cooled supports. Thermal properties of selected materials are available (173).

Ion Implantation Economics

Directed Beam Ion Implantation. The costs of using directed beam implantation have been estimated (2) by assuming set expenses for capital equipment, maintenance, staff, number of shifts (1, 2, or 3), utilities, and consumables by relating these expenses to predicted throughputs (cm² h) for both spherical and flat targets. Implanting spheres to a dose of 2 × 10¹⁷ ions cm² using a 10 mA beam and one working shift (the most expensive case) is estimated to cost \$0.64 cm². The least expensive scenario (\$0.11 cm²) implanted flat targets under similar conditions, but used three shifts. The principal difference between implanting spheres or flats is that spheres require target manipulation and masking which decrease the beam utilization and therefore increase the cost. In addition, the study concluded that beam heating would constrain the beam current density and would require a high throughput facility to use a large diffuse beam. The estimated costs of implantation were compared to the estimated costs of part replacement and, depending on the cost of the part, cost savings up to factors of five were predicted.

The costs of metal ion implantation using the MEVVA technology have also been estimated (3). Significant capital equipment savings are claimed over traditional directed beam systems (as used in the semiconductor industry) because expensive mass analyzing magnets are not required by the MEVVA design. The cost estimates have the advantage over other estimates in that the authors have built and sold commercial systems and therefore base the capital equipment costs on the net present value of an operation system. For a three-shift operation implanting 8000 cm², the cost is estimated to be \$0.13 cm². Only for coming generations of MEVVA systems do projected costs approach \$0.01 cm². Interesting comparisons to other coating technologies (especially physical vapor deposition, PVD) are also made (3).

Omnidirectional Ion Implantation. Adopting the methodology of Reference r2 to include machine availability time, the costs of using plasma source ion implantation (PSII) have been predicted (4). The costs of capital equipment, consumables, utilities, and rented space were rigorously accounted for. The implantation costs were computed as a function of throughput (m² yr) and the time to process one batch of targets (h). The use of nitrogen gas, an implanted dose of 2 × 10¹⁷ ions cm², and the use of two working shifts were assumed. If a PSII facility can implant 1000 m² year, the cost is estimated to be \$0.10 cm² if the processing time is less than eight hours. However, if the facility implants 10,000 m² year, the costs could vary from \$0.01 cm² for a processing time of one hour to about \$0.04 cm² for a processing time of eight hours. Since an omnidirectional implantation process is better suited for nonplanar targets, it was concluded that the PSII process can offer substantial savings over directed beam implantation processes (4).

BIBLIOGRAPHY

"Ion Implantation" in *ECT* 3rd ed., Vol. 13, pp. 706–719, by J. W. Butler, Naval Research Laboratory.

1. J. W. Mayer, L. Eriksson, and J. A. Davies, *Ion Implantation of Semiconductors*, Academic Press, Inc., New York, 1970.
2. F. A. Smidt and B. D. Sartwell, *Nucl. Instr. Meth. Phys. Res.* **B6**, 70 (1985).
3. J. R. Treglio, A. J. Perry, and R. J. Skinner, *Surf. and Coatings Technol.* **65**, 184 (1994).
4. D. J. Rej and R. B. Alexander, *J. Vac. Sci. Tech. B* (Aug. 1994). Also available as a Los Alamos National Laboratory, preprint LA-UR #94–312, Los Alamos, N.M., 1994.
5. G. A. Collins and co-workers, in *Proceedings of the International Conference on Metallurgical Coatings and Thin Films*, Apr. 25–29, 1994, San Diego, Calif.
6. J. F. Ziegler, *Handbook of Stopping Cross-Sections for Energetic Ions in All Elements*, Pergamon Press, New York, 1980.
7. J. W. Mayer and S. S. Lau, *Electronic Materials Science: For Integrated Circuits in Si and GaAs*, Macmillian, New York, 1990, Chapt. 8.
8. J. Lindhard, M. Scharff, and H. E. Shiott, *Mat. Fys. Medd. Dan. Vid. Selsk.* **33**(14), 3 (1963).
9. J. P. Biersack and L. G. Haggmark, *Nucl. Instr. and Methods* **174**, 257–269 (1980).
10. J. F. Ziegler, J. P. Biersack, and U. Littmark, *The Stopping and Range of Ions in Solids*, Pergamon Press, Inc., New York, 1985.
11. K. B. Winterbon, P. Sigmund, and J. B. Sanders, *Mat. Fys. Medd. Dam. Vid. Selsk.* **37**(14), 5 (1970).
12. Y. Kido and J. Kawamoto, *App. Phys. Lett.* **48**, 257 (1986).
13. R. Simonton and A. F. Tasch, in J. F. Ziegler, ed., *Handbook of Ion Implantation Technology*, North-Holland, New York, 1992, p. 119.
14. M. T. Robinson, *Phil. Mag.* **12**, 741 (1965).
15. M. T. Robinson and O. S. Oen, *J. Nucl. Mater.* **110**, 147 (1982).
16. P. Sigmund, *Radiat. Eff.* **1**, 15 (1969).
17. G. H. Kinchin and R. S. Pease, *Rept. Prog. Phys.* **18**, 1 (1955).
18. M. W. Thompson, *Defects and Radiation Damage in Metals*, Cambridge University Press, New York 1969.
19. D. R. Orlander, *Fundamental Aspects of Nuclear Reactor Fuel Elements*, National Technical Information Service, Springfield, Va., 1976, Chapt. 17.
20. M. J. Norget, M. T. Robinson, and I. M. Torrens, *Nucl. Eng. Des.* **33**, 50 (1975).
21. N. Andersen and P. Sigmund, *Mat. Fys. Medd. Dan. Vid. Selsk.* **39**(3), 1 (1974).
22. C. A. Coulter and D. M. Parkin, *J. Nucl. Mater.* **88**, 249 (1980).

23. D. M. Parkin and C. A. Coulter, *J. Nucl. Mater.* **101**, 261 (1981).
24. D. M. Parkin, in C. J. McHargue, R. Kossowsky, and W. O. Hofer, eds., *Structure–Property Relationships in Surface-Modified Ceramics*, Kluwer Academic Publishers, Dordrecht, 1989, p. 47.
25. Y. Adda, M. Beyeler, and G. Brebec, *Thin Solid Films* **25**, 107 (1975).
26. R. Sizmann, *J. Nucl. Mater.* **69/70**, 386 (1978).
27. K. C. Russell, *Progr. Mater. Sci.* **28**, 229 (1984).
28. H. Hseih, T. Dias del la Rubia, and R. S. Averback, *Phys. Rev.* **B40**, 9986 (1989).
29. M. Nastasi and J. W. Mayer, *Radiation Effects and Defects in Solids* **130–131**, 367 (1994).
30. N. Matsunami and co-workers, *Atomic Data and Nuclear Data Tables* **31**, 1–84 (1984).
31. P. Sigmund, *Benirsch* **81**, 9 (1981).
32. C. Kittel, *Introduction to Solid State Physics*, 6th ed., John Wiley & Sons, Inc., New York, 1976, Chapt. 3.
33. Y. Yamamura and N. Itoh, in T. Itoh, ed., *Ion Beam Assisted Film Growth*, Elsevier, Amsterdam, 1989, Chapt. 4.
34. K. S. Grabowski and co-workers, *Mater. Res. Soc. Symp. Proc.* **27**, 615 (1984).
35. G. K. Hubler, *NRL Memorandum Report 5928*, Naval Research Laboratory, Washington, D.C., Mar. 13, 1987.
36. Z. L. Liao and J. W. Mayer, in J. K. Hirvonen, ed., *Ion Implantation*, Academic Press, Inc., New York, 1980, Chapt. 2.
37. G. Betz and G. K. Wehner, *Benirsch* **83**, 11 (1983).
38. H. H. Anderson, in J. S. Williams and J. M. Poate, eds., *Ion Implantation and Beam Processing*, Academic Press, Inc., New York, 1980, Chapt. 6.
39. L. Clapham, J. L. Whitton, and M. C. Ridgway, *Mat. Lett.* **16**, 139 (1993).
40. B. Hoffmann and co-workers, *Nucl. Instr. Methods Phys. Res.* **B36**, 30 (1989).
41. D. A. Baldwin, B. D. Sartwell, and I. L. Singer, *Nucl. Instr. Methods Phys. Res.* **B7–8**, 49 (1985).
42. B. D. Sartwell and D. A. Baldwin, *Mater. Sci. Eng.* **69**, 539 (1985).
43. W. Eckstein, *Computer Simulation of Ion–Solid Interactions*, Springer-Verlag, Berlin, 1991.
44. M. T. Robinson and I. M. Torrens, *Phys. Rev.* **B9**, 5008 (1974); M. T. Robinson, *Phys. Rev.* **B67**, 396 (1992).
45. M. T. Robinson, *Nucl. Instr. Methods Phys. Res.* **B48**, 408 (1990).
46. M. T. Robinson, *Phys. Rev.* **B40**, 10717 (1989).
47. J. M. Haile, *Molecular Dynamics Simulations*, Wiley-Interscience, New York, 1992.
48. T. Diaz de la Rubia and M. W. Guinan, in M. van Rossum, ed., *Trends in Ion Implantation*, Trans Tech. Publications, Aedermannsdorf, Switzerland, 1992.
49. A. F. Voter, in J. H. Westbrook and R. L. Fleischer, eds., *Intermetallic Compounds: Principles and Practice*, John Wiley & Sons, Inc., New York, 1993.
50. W. F. van der Weg, D. Sigurd, and J. W. Mayer, in S. T. Picraux, E. P. EerNisse, and F. L. Vook, eds., *Applications of Ion Beams to Metals*, Plenum Press, New York, 1974, p. 209.
51. J. W. Mayer and S. S. Lau, in J. M. Poate, G. Foti, and G. J. Clark, eds., *Surface Modification and Alloying by Laser, Ion, and Electron Beams*, Plenum Press, New York, 1983, p. 241.
52. L. E. Rehn and P. R. Okamoto, *Nucl. Instr. Meth.* **B39**, 104 (1989).
53. R. S. Averback, *Nucl. Instr. Meth.* **B15**, 675 (1986).
54. Y.-T. Cheng, *Mat. Sci. Rep.* **5**, 45 (1990).
55. M. Nastasi and J. W. Mayer, *Mater. Sci. Eng.* **R1**(1), 1–49 (1994).
56. H. A. Atwater, *Solid State Phenomena* **27**, 67 (1992).
57. J. J. Cuomo, S. M. Rossnagel, and H. R. Kaufman, eds., *Handbook of Ion Beam Processing Technology*, Noyes Publications, Park Ridge, N. J., 1989.
58. T. Itoh, ed., *Ion Beam Assisted Film Growth*, Elsevier, Amsterdam, 1989.
59. O. Auciello and R. Kelly, eds., *Ion Bombardment Modification of Surfaces: Fundamentals and Applications*, Elsevier, Amsterdam, 1984.
60. J. Harper, in *Plasma Surface Interactions and Processing of Materials*, Kluwer Academic Publishers, the Netherlands, 1990, p. 251.
61. N. Herbots and co-workers, in J. W. Rabalais, ed., *Low Energy Ion Surface Interaction*, John Wiley & Sons, Inc., New York, 1994.
62. J. K. Hirvonen, *Mater. Sci. Rep.* **6**, 215 (1991).
63. F. A. Smidt, *Internat. Mat. Rev.* **35**, 61 (1990).
64. S. Aisenberg and R. Chabot, *J. Appl. Phys.* **42**, 2953 (1971).
65. C. Weissmantel, *J. Vac. Sci. Technol.* **18**, 179 (1981).
66. M. Kitabatake and K. Wasa, in Ref. 57.
67. J. W. Rabalais and S. Kasi, *Science* **239**, 623 (1989).
68. Y. Andoh and co-workers, *Nucl. Instr. Methods Phys. Res.* **B19/20**, 787 (1987).
69. S. Nishiyama and co-workers, *Nucl. Instr. Methods Phys. Res.* **B80/81**, 1485 (1993).
70. D. Kester and R. Messier, *J. Appl. Phys.* **72**(2), 504 (1992).
71. M. Kuichi and co-workers, *Nucl. Instr. Methods Phys. Res.* **B37/38**, 701 (1989).
72. B. O. Johansson and co-workers, *J. Mat. Res.* **1**, 442 (1986).
73. P. J. Martin and co-workers, *Appl. Optics* **22**, 178 (1983).
74. P. J. Martin, R. P. Netterfield, and D. R. McKenzie, *Thin Solid Films* **137**, 207 (1986).
75. W. T. Pawlewicz and co-workers, *SPIE Proc.* **1618**, 1 (1991).
76. E. P. Donovan and co-workers, *Appl. Optics* **28**, 2940 (1989).
77. O. Marcovitch, Z. Klein, and I. Lubezky, *Appl. Optics* **28**, 2792 (1989).
78. G. A. Al-Jumaily and co-workers, *J. Vac. Sci. Technol.* **A5**, 1909 (1987).
79. F. Case, *J. Vac. Sci. Technol.* **5**, 1762 (1987).
80. M. Nagakubo, T. Yamamoto, and M. Naoe, *Mater. Res. Soc. Symp. Proc.* **128**, 29 (1989).
81. R. A. Kant and B. D. Sartwell, *J. Vac. Sci. Technol.* **A8**, 86 (1990).
82. G. K. Hubler and co-workers, *Mater. Res. Soc. Symp. Proc.* **128**, 55 (1989).
83. T. Miyano and H. Kitamura, *Proceedings 8th International Conference on Surface Modification of Metals by Ion Beams*, Sept. 1993, Kanazawa, Japan.
84. M. Barth and co-workers, *Proceedings of 3rd International Conference on Surface Modification Technologies*, Neuchatel, Switzerland, 1989.
85. Y. Andoh and co-workers, *Nucl. Instr. Methods in Phys. Res.* **B80/81**, 225 (1993).
86. H. Kuwano and K. Nagai, *J. Vac. Sci. Technol.* **A4**, 2993 (1986).
87. A. Erdemer and co-workers, STLE Preprint No. 89-AM-5C-1, 1989.
88. E. McCafferty and co-workers, *Mater. Sci. Eng.* **86**, 1 (1987).
89. W. Ensinger and G. K. Wolf, *Mater. Sci. and Eng.* **A116**, 1 (1989).
90. M. A. Abuzriba and co-workers, *Corrosion* **48**, 2 (1992).
91. N. E. W. Hartley, *Tribology* **8**, 65 (1975).
92. C. A. Carosella and co-workers, in Preece and Hirvonen, eds., *Ion Implantation Metallurgy*, AIME, Warrendale, Pa., 1980, p. 321.
93. J. K. Hirvonen, in L. E. Rehn, S. T. Picraux, and H. Wiedersich, eds., *Surface Alloying by Ion, Electron, and Laser Beams*, American Society of Metals (ASM), Metals Park, Ohio, 1987, p. 373.
94. F. Alonso and co-workers, *Nucl. Instrm. Methods Phys. Res.* **B80/81** (1993).
95. J.-P. Hirvonen and co-workers, *Mat. Res. Soc. Symp. Proc.* **316** (1994).

96. B. S. Covino and co-workers, *J. Electrochem. Soc.* **125**, 366 (1978).
97. K. Takahashi and co-workers, *Nucl. Instrm. Methods Phys. Res.* **182/183**, 1009 (1981).
98. J. M. Williams and co-workers, in Hubler and co-workers, eds., *Ion Implantation and Ion Beam Processing of Materials*, Elsevier, New York, 1984, p. 735.
99. G. Deamaley, in B. M. Ulrich, ed., *Proceedings 4th International Conference on Ion Beam Modification of Materials*, North Holland, Amsterdam, p. 158.
100. P. Sioshansi, *Nucl. Instrm. Methods Phys. Res.* **B24/25**, 767 (1987).
101. J. K. Hirvonen, *Mater. Sci. Eng.* **A116**, 167 (1989).
102. N. J. Mikkelsen and C. A. Straede, *Surf. and Coatings Technol.* **51**, 152 (1992).
103. B. Bhushan and B. K. Gupta, *Handbook of Tribology: Materials, Coatings, and Surface Treatments*, McGraw-Hill Book Co. Inc., New York, 1991; P. J. Blau, *Friction, Lubrication, and Wear Technology*, Vol. 18 *American Society of Metals (ASM) Handbook Series*, ASM International, Metals Park, Ohio, 1992.
104. M. Iwaki, *Crit. Rev. Solid State Mater. Sci.* **15**(5), 473 (1989).
105. J. K. Hirvonen, *Surf. Coatings Technol.* **65**, 84 (1994).
106. R. J. Culbertson and co-workers, *Nucl. Instrm. Methods Phys. Res.* **56/57**, 652 (1991).
107. A. Kujore and co-workers, *Nucl. Instrm. Methods Phys. Res.* **182/183**, 949 (1981).
108. K. Jata and E. A. Starke, Jr., *J. Met.* **35**(8), 23 (1983).
109. W. W. Hu and co-workers, in C. M. Preece and J. K. Hirvonen, eds., *Ion Implantation Metallurgy*, TMS-AIME, Warrendale, Pa., 1980, p. 92.
110. R. G. Vardiman and J. E. Kox, *Acta Metall.* **33**, 2033 (1985).
111. R. G. Vardiman and R. A. Kant, *J. Appl. Phys.* **53**, 690 (1982).
112. K. V. Jata and co-workers, *Scripta Met.* **17**, 479 (1983).
113. G. Deamaley, in V. Ashworth, W. A. Grant, and R. P. M. Procter, eds., *Ion Implantation into Metals*, Pergamon Press, New York, 1982, p. 180.
114. G. K. Hubler, in S. T. Picraux and W. J. Choyke, eds., *Metastable Materials Formation by Ion Implantation*, North Holland, Amsterdam, 1982.
115. G. Deamaley, *IEEE Trans. on Nucl. Sci.* **NS-28**, 1808 (1981).
116. R. G. Vardiman and co-workers, *Materials Evaluation under Fretting Conditions*, ASTM-STP780, ASTM, 1982, p. 138.
117. R. G. Vardiman, in R. F. Hochman, ed., *Ion Plating and Implantation*, ASM International, Metals Park, Ohio, 1986, p. 107.
118. M. J. Bennett and A. T. Tuscon, *Mater. Sci. Eng.* **A116**, 79 (1989).
119. C. R. Clayton, in L. E. Rehn, S. T. Picraux, and H. Wiedersich, eds., *Surface Alloying by Ion, Electron, and Laser Beams*, American Society of Metals (ASM), Metals Park, Ohio, 1991, p. 325.
120. K. M. Kramer, J. R. Tesmer, and M. Nastasi, *Nucl. Instr. Methods Phys. Res.* **B59/60**, 860 (1991).
121. V. Ashworth, R. P. M. Procter, and W. A. Grant, in Ref. 36.
122. E. McCafferty and G. K. Hubler, *J. Electrochem. Soc.* **128**, 1892 (1978).
123. P. Munn and G. K. Wolf, *Mater. Sci. Eng.* **69**, 303 (1985).
124. C. R. Clayton and co-workers, in M. Froment, ed., *Passivity of Metals and Semiconductors*, Elsevier, Amsterdam, 1983, p. 305.
125. R. Valori and G. K. Hubler, *NRL Memorandum Report 4527*, 99 (1981).
126. J. L. Whitton, W. A. Grant, and J. L. Williams, *Proceedings International Conference on Ion Beam Modification of Materials*, Budapest, 1981.
127. F. A. Smidt and B. D. Sartwell, *Mat. Sci. and Eng.* **90**, 1 (1987).
128. Y. F. Wang and co-workers, *Thin Solid Films* **63**, 11 (1979).
129. J. A. Cairns, *Applications of Low Energy Accelerators*, Institute of Electrical & Electronics Engineers, Denton, Tex. 1980.
130. P. Rabette and co-workers, *Chem. Phys. Letters* **60**, S.348 (1979).
131. W. E. O'Grady and G. K. Wolf, *Proceedings Electrochemical Society*, Minneapolis, May 1981.
132. G. K. Wolf, *Chemie Ingenieur Technik*, (1982).
133. H. Kasten and G. K. Wolf, *Electrochem. Act.* **25**, 1581 (1980).
134. T. Hioki and co-workers, in C. McHargue and co-workers, eds., *Structural-Property Relationships in Surface Modified Ceramics*, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1989, p. 303.
135. P. J. Burnett and T. F. Page, *J. Mat. Sci.* **19**, 3524 (1984).
136. C. J. McHargue and C. S. Yust, *J. Amer. Ceram. Soc.* **67**(2), 117 (1984).
137. D. W. Oblas and V. K. Sarin, in R. F. Hochman, H. S. Legg, and K. O. Legg, eds., *Ion Implantation and Plasma Assisted Processes*, American Society of Metals, Metals Park, Ohio, 1988, p. 95.
138. C. W. White, C. J. McHargue, and P. S. Sklad, *Mat. Sci. Rep.* **4**(2 3), 41 (1989).
139. R. Reeber, *J. Amer. Ceram. Soc.* **76**(2), 261 (1993).
140. V. Venkatesan, *Nucl. Instrm. Methods Phys. Res.* **B7/8**, 461 (1985).
141. E. H. Lee and co-workers, *Nucl. Instrm. Methods Phys. Res.* **B74**, 326 (1993).
142. B. Banks, in O. Auciello and R. Kelly, eds., *Surface Modification of Surfaces*, Elsevier, New York, 1984; S. A. Jenekhe, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, 2nd ed., John Wiley & Sons, Inc., New York, 1989, p. 352.
143. M. Nastasi and J. W. Mayer, *Mater. Sci. Rep.* **6**, 1 (1991).
144. P. R. Okamoto and M. Meshii, in H. W. Wiedersich and M. Meshii, eds., *Science of Advanced Materials*, American Society of Metals, Metals Park, Ohio, 1990, p. 33.
145. S. Aisenberg and R. Chabot, *J. Appl. Phys.* **42**, 2953 (1971).
146. R. Lossy and co-workers, *Appl. Phys. Lett.* **61**(2), 171 (1992).
147. Y. Lifshitz, S. R. Kasi, and J. W. Rabalais, *Phys. Rev. Lett.* **62**, 1290 (1989).
148. Y. Lifshitz, S. R. Kasi, and J. W. Rabalais, *Phys. Rev. Lett.* **B41**, 10,468 (1990).
149. J. Robertson, *Prog. Solid. St. Chem.* **21**, 199 (1991).
150. M. W. Geis and M. A. Tamor, in G. L. Trigg, ed., *Encyclopedia of Applied Physics*, VCH Publishers, New York, Vol. 5, 1993.
151. *Status and Applications of Diamond and Diamond-Like Materials: An Emerging Technology*, National Academy Press, Washington, D.C., 1990.
152. J. C. Angus, P. Koidl, and S. Domitz, in J. Mort and F. Jansen, eds., *Plasma Deposited Thin Films*, CRC Press, Inc., Boca Raton, Fla., 1986, p. 89.
153. J. S. Williams, *Rep. Prog. Phys.* **49**, 491 (1986).
154. W. D. Cussins, *Proc. Phys. Soc.* **B68**, 213 (1955).
155. P. Rose, *Nucl. Instrm. Methods Phys. Res.* **B6**, 1 (1985).
156. I. G. Brown, ed., *The Physics and Technology of Ion Sources*, John Wiley & Sons, Inc., New York, 1989.
157. G. D. Alton, *Nucl. Instrm. Methods Phys. Res.* **189**, 15 (1981).
158. H. Freeman, in G. Deamaley, eds., *Ion Implantation*, North Holland, Amsterdam, 1972.
159. R. Keller, in I. G. Brown, ed., *The Physics and Technology of Ion Sources*, John Wiley & Sons, Inc., New York, 1989.
160. B. Torp and co-workers, *Rev. Sci. Instruments* **61**, 595 (1990).
161. I. Brown, *J. Vac. Sci. Technol.* **A11**(4), 1480 (1993).
162. J. R. Treglio, G. D. Magnuson, and R. J. Skinner, *Surf. Coatings Technol.* **51**, 546 (1992).
163. N. Sakudo and co-workers, *Rev. Sci. Instr.* **48**, 762 (1977).
164. H. Ryssel and H. Glawischung, eds., *Ion Implantation Techniques*, Springer-Verlag, New York, 1982.
165. A. Wittkower and J. K. Hirvonen, *Nucl. Instrm. Methods Phys. Res.* **B6**, 78 (1985).
166. M. M. Shamim and co-workers, *J. Appl. Phys.* **70**, 4756 (1991).
167. J. R. Conrad and co-workers, *J. Appl. Phys.* **62**, 4591 (1987).

168. U.S. Pat. **4,764,394** (Aug. 16, 1988), J. R. Conrad (to Wisconsin Alumni Research Foundation).
169. I. Brown and J. Washburn, *Nucl. Instrm. Methods Phys. Res.* **B21**, 201 (1987).
170. W. A. Reass, *J. Vac. Sci. Technol.* **B12**, 854 (1994).
171. B. P. Wood and co-workers, *Mat. Res. Soc. Symp. Proc.* **279**, 345 (1993).
172. X. Y. Qian and co-workers, *Nucl. Instrm. Methods Phys. Res.* **B55**, 84 (1991).
173. K. Grabowski and R. A. Kant, in H. Ryssel and H. Glawischnig, eds., *Ion Implantation: Equipment and Techniques*, Springer-Verlag, New York, 1983, p. 364.

General References

Ion–Solid Interactions] F. Ziegler, J. P. Biersack, and U. Littmark, *The Stopping and Range of Ions in Solids*, Pergamon Press, Inc., New York, 1985.
M. A. Kunakhov and F. F. Komarov, *Energy Loss and Ion Ranges in Solids*, Gordon and Breach Science Publishers, New York, 1981.
J. W. Mayer, L. Eriksson, and J. A. Davies, *Ion Implantation in Semiconductors*, Academic Press, Inc., New York, 1970.
Z. L. Liao and J. W. Mayer, "Ion Bombardment Effects on Material Composition," in J. K. Hirvonen, ed., *Ion Implantation*, Academic Press, Inc., New York, 1980, Chapt. 2.
H. H. Anderson, "Ion-Bombardment-Induced Composition Changes in Alloys and Compounds," in J. S. Williams and J. M. Poate, eds., *Ion Implantation and Beam Processing*, Academic Press, Inc., New York, 1984, Chapt. 6.
R. Behrisch, ed., "Sputtering by Particle Bombardment I: Physical Sputtering of Single Element Solids," in *Topics in Applied Physics*, Vol. 47, Springer-Verlag, Berlin, 1981.
R. Behrisch, ed., "Sputtering by Particle Bombardment II: Sputtering of Alloys and Compound, Electron and Neutron Sputtering, Surface Topography," in *Topics in Applied Physics*, Vol. 52, Springer-Verlag, Berlin, 1983.
R. Behrisch, R. Wittmaack, and K. Wittmaack, eds., "Sputtering by Particle Bombardment, III: Characteristics of Sputtered Particles, Technical Applications," in *Topics in Applied Physics*, Vol. 64, Springer-Verlag, Berlin, 1991.
G. Deamaley, J. H. Freeman, R. S. Nelson, and J. Stephen, *Ion Implantation*, North Holland, Amsterdam, 1973.
R. S. Averback, "Ion-Beam Modification of Materials," *Encyclopedia of Applied Physics*, VCH Publishers, New York, Vol. 8, 1994, p. 173.
M. Nastasi, J. W. Mayer, and J. K. Hirvonen, *Ion Solid Interactions: Fundamentals and Applications*, Cambridge University Press, Cambridge, in press.

Ion Implantation Applications

J. R. Conrad and K. Sridharan, eds., "Proceedings of the 1st International Workshop on Plasma Based Ion Implantation," *J. Vac. Sci. Tech.* **B12**, 807 (1994).
S. M. Rossnagel and J. J. Cuomo, "Ion-Beam-Assisted-Deposition and Synthesis," *MRS Bulletin*, 40 (Feb. Mar. 1987).
J. F. Ziegler, ed., *Handbook of Ion Implantation Technology*, North-Holland, New York, 1992.
G. K. Hubler, *Ion Beam Processing*, NRL Memorandum Report 5928, Mar. 13, 1987.
B. Bhushan, and B. K. Gupta, *Handbook of Tribology: Materials, Coatings, and Surface Treatments*, McGraw-Hill Book Co., Inc., New York, 1991.
R. A. Buchanan and E. E. Stansbury, "Aqueous Corrosion", in R. Kossowsky, ed., *Surface Modification Engineering: Fundamental Aspects*, CRC Press, Boca Raton, Fla., 1989.
J. F. Ziegler, ed., *Ion Implantation Technology*, North-Holland, Amsterdam, 1992.
I. G. Brown, ed., *The Physics and Technology of Ion Sources*, John Wiley & Sons, Inc., New York, 1989.
J. F. Ziegler, ed., *Ion Implantation: Science and Technology*, Academic Press, Inc., New York, 1984.
D. M. Follstaedt, "Ion Implantation and Ion-Beam Mixing," in R. W. Cahn, P. Haasen, and E. J. Kramer, eds., *Materials Science and Technology*, Vol. 15.
H. Ryssel and H. Glawischung, eds., *Ion Implantation Techniques*, Springer-Verlag, Berlin, 1982.
H. R. Kaufman and R. S. Robinson, *Operation of Broad-Beam Sources*, Commonwealth Scientific, Alexandria, Va., 1987.
J. W. Mayer and S. S. Lau, *Electronic Materials Science: For Integrated Circuits in Si and GaAs*, Macmillian, New York, 1990.

Ion Implantation Technology

J. F. Ziegler, ed., *Ion Implantation Technology*, North-Holland, Amsterdam, 1992.
I. G. Brown, ed., *The Physics and Technology of Ion Sources*, John Wiley & Sons, Inc., 1989.
J. F. Ziegler, ed., *Ion Implantation: Science and Technology*, Academic Press, Inc., New York, 1984.
H. Ryssel and H. Glawischung, eds., *Ion Implantation Techniques*, Springer-Verlag, Berlin, 1982.
H. R. Kaufman and R. S. Robinson, eds., *Operation of Broad-Beam Sources*, Commonwealth Scientific, Alexandria, Va., 1987.
G. Deamaley, "Historical Perspective of Ion Implantation," in *Proceedings 8th International Conference on Surface Modification of Metals by Ion Beams, Kanazawa, Japan*, North-Holland, Amsterdam.
J. K. Hirvonen, "Ion Beam Processing for Industrial Applications," *Mater. Sci. Eng.* **A116**, 167 (1992).

Kevin C. Walter
Michael Nastasi
Los Alamos National Laboratory

IONOMERS

The generic term ionomer was introduced by Du Pont in 1964 (1) in conjunction with the commercialization of the new Surlyn resins to denote a thermoplastic polymer containing both covalent and ionic bonds, and having properties influenced to substantial effect by the ionic bonding. Since that time, the meaning has been expanded to include many compositions such as the glass ionomers used in dentistry (2) which cannot be melt processed (see DENTAL MATERIALS). In the interest of clarity and consistency, it is proposed that the term ionomer be reserved for polymers having melt viscosities suitable for conventional melt processing methods. Descriptions such as ion-containing or ion-linked are appropriate for highly viscous or true thermoset materials.

Within the scope of the original definition, a very wide variety of ionomers can be obtained by the introduction of acidic groups at molar concentrations below 10⁰ % into the important addition polymer families, followed by partial neutralization with metal cations or amines. Extensive studies have been reported, and useful reviews of the polymers have appeared (3–8). Despite the broad scope of the field and the unusual property combinations obtainable, commercial exploitation has been confined mainly to the original family based on ethylene copolymers. The reasons for this situation have been discussed (9). Within certain industries, such as flexible packaging, the word ionomer is understood to mean a copolymer of ethylene with methacrylic or acrylic acid, partly neutralized with sodium or zinc.

This article focuses on the commercial, ethylene-based ionomers and includes information on industrial uses and manufacture. The fluorinated polymers used as membranes are frequently included in ionomer reviews. Owing to the high concentration of polar groups, these polymers are generally not melt processible and are specially designed for specific membrane uses (see FLUORINE COMPOUNDS, ORGANIC–PERFLUOROALKANE SULFONIC ACIDS; MEMBRANE TECHNOLOGY).

Ethylene-Based Ionomers

Ionomer resins consisting of ethylene–methacrylic acid copolymers partially neutralized with sodium or zinc were commercially introduced in 1964 by Du Pont under the Surlyn trademark (1). More recently, a similar line of products, sold as Hi-Milan resins, has been commercialized by Mitsui–Du Pont in Japan. Iolon ionomeric resins, based on ethylene–acrylic acid, are produced by Exxon in Belgium. Ionomers containing about 1 mol % of carboxylate groups are offered by BP in Europe as Novex resins. Low molecular weight, waxy Adlyn ionomers are produced and sold by AlliedSignal.

PHYSICAL PROPERTIES

The semicrystalline, ethylene-based ionomers of commerce are flexible, transparent polymers notable for high strength and elasticity in both solid and molten states. The ionic bonding is completely reversible (8) and has a strong influence on properties, even at temperatures well above the melting point.

Table 1. Mechanical Properties of Surlyn Ionomers^a

Property	Range	ASTM test
stiffness, MPa ^b	90–400	D747 and D882
yield point, MPa ^b	8–20	D412
tensile strength, MPa ^b	23–40	D638
elongation at break, %	280–500	D412
Shore D	54–70	D785
brittleness temperature, °C	100 to 140	D746

^a Ref. 6.
^b To convert MPa to psi, multiply by 145.

Mechanical Properties. Table 1 shows the general range of mechanical properties available in commercial Surlyn ionomers (6). The range can be extended in the direction of lower stiffness for special applications by incorporating termonomers, which decrease crystallinity (10). In the first descriptions (11) of the relationships between mechanical properties and content of ionized groups, it was shown that stiffness, yield point, and tensile strength all increase with increasing acid content and degree of neutralization, whereas the elongation at break decreases moderately. A comparison between the stress–strain curves of an ionomer and a conventional branched polyethylene (Fig. 1) shows that the stress does not level off for ionomers above the yield point, resulting in very high energy to break. The shape of the stress–strain curve also accounts for the unusual resilience of ionomers, which is critical for uses in sporting goods and footwear.

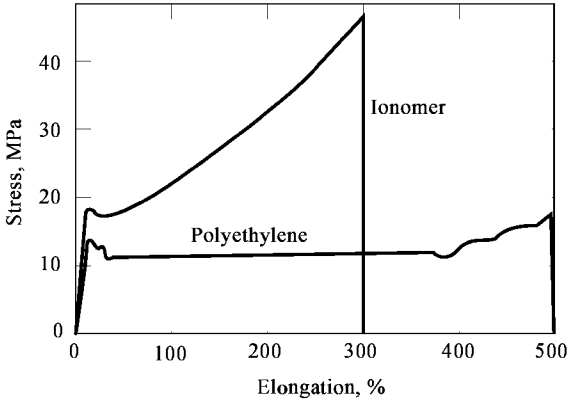


Fig. 1. Stress–strain curves for ionomer and polyethylene resins. Test speed is 5 cm min. The reference material is high molecular weight conventional polyethylene, density 0.920 (11). To convert MPa to psi, multiply by 145.

Mechanical and melt-flow data for a variety of ionomers, all based on an ethylene–methacrylic acid copolymer of the type used in commercial practice, are shown in Table 2 (12). Seven metal-ion types have been used and the similarities among the polymers containing ions from many groups of the Periodic Table are readily apparent. However, there are also interesting anomalies, such as the very low melt flow of the magnesium ionomers. Certain other metals, notably aluminum, give products which are not readily melt processible (8). Table 2 includes data on the effect of room temperature aging on the stiffness of ionomers. The tendency to become stiffer over long time periods is more pronounced in ionomers than in other crystalline polymers. Data for the 60° neutralized zinc ionomer show the extent of this change in modulus.

Table 2. Rheological and Physical Properties of EMAA Copolymers ^{a, b}

Sample designation ^b	Melt flow rate, ^c dg min	Stiffness, MPa ^d			Tensile properties		
		Aging durations			Yield stress, MPa ^d	Tensile strength, MPa ^d	Elongation at break, °
		9 d	2 mo	3 mo			
EMAA	60	70	82	83	8.2	22.0	530
EMAA,0.2 Na	18		169	196	12.9	24.6	473
EMAA,0.4 Na	4.6		277	320	16.1	30.4	387
EMAA,0.6 Na	1.0		256	325	16.0	31.5	295
EMAA,0.8 Na	0.23		245	310	18.0	34.3	324
EMAA,0.9 Na	0.18		248	309	17.2	34.0	319
EMAA,0.2 K	17		137	167	13.0	24.5	462
EMAA,0.4 K	4.7		236	238	15.3	24.4	320
EMAA,0.6 K	1.3		245	231	14.8	28.1	417
EMAA,0.2 Mg	18		191	252	14.9	25.9	439
EMAA,0.4 Mg	4.5		285	362	18.8	29.7	350
EMAA,0.6 Mg	0.80		254	321	19.2	29.0	299
EMAA,0.8 Mg	0.06		208	276	21.3	28.0	196
EMAA,0.9 Mg	0.06		205	280	20.4	21.5	72
EMAA,0.2 Zn	23	109	124	140	10.9	22.2	412
EMAA,0.4 Zn	8.7		183	235	15.8	24.9	400
EMAA,0.6 Zn	3.6	223	255	302	15.4	27.8	375
EMAA,0.8 Zn	1.0		298	343	17.9	24.2	267
EMAA,0.9 Zn	0.40		298	347	20.1	22.7	126
EMAA,0.2 Cu	23	124			10.3	22.1	452
EMAA,0.4 Cu	9.9	186			12.9	21.3	408
EMAA,0.6 Cu	5.7	243			15.7	17.6	208
EMAA,0.2 Mn	31	123					
EMAA,0.4 Mn	18	196			13.3	27.4	433
EMAA,0.6 Mn	10	278			15.8	28.2	375
EMAA,0.2 Co	16	160			21.8	24.6	437
EMAA,0.4 Co	9.8	249			16.7	17.8	121
EMAA,0.6 Co	5.2	288			17.4	17.5	95

^a Ref. 12.
^b EMAA = ethylene–methacrylic acid. Decimal numbers are decimal fractions of acid groups that have been neutralized.
^c At 190°C.
^d To convert MPa to psi, multiply by 145.

Similar mechanical data for a series of ionomers derived from a single ethylene–acrylic acid copolymer have appeared (13) (Table 3). Comparison of the data from Tables 2 and 3 shows that the substitution of acrylic acid for methacrylic acid has only minor effects on properties.

Table 3. Properties of Poly(ethylene-*co*-acrylic acid) Salts ^a

Material	Percent neutralized	Melt index at 190°C, dg min	Secant modulus, ° extension, MPa ^b	Ultimate tensile strength, MPa ^b	Elongation at break, °
control material, 14.8 wt ° acrylic acid	0	67	48.26	14.8	470
sodium salt ^c	12.0	12.2	230	21.7	420
	30.0	3.9	540	27.6	330
	47.5	1.0	293	31.7	310
	66.0	0.3	273	33.0	280
potassium salt ^d	8.0	16.3	202	21.0	470
	25.0	4.5	362	25.5	410
	51.0	2.7	341	30.6	370
	63.0	0.57	308	34.4	390
lithium salt ^e	12.0	18	181	21.7	410
	28.5	5.2	337	26.5	350
	52.5	1.4	334	28.2	260
	67.5	0.2	254	31.7	250
acrylic acid, polymer with ethylene, calcium salt	11.03063.5	16.83.10.08	180270280	202531	465390200
acrylic acid, polymer with ethylene, magnesium salt	15.02537.5	198.52.5	185251278	17.52223.5	390385330

^a Ref. 13.
^b To convert MPa to psi, multiply by 145.
^c Acrylic acid, polymer with ethylene, sodium salt [25750-82-7], C₃H₄O₂·C₂H_{4x}·xNa.

^d Acrylic acid, polymer with ethylene, potassium salt [27515-34-0], (C₃H₄O₂·C₂H₄)_x·xK.
^e Acrylic acid, polymer with ethylene, lithium salt, (C₃H₄O₂·C₂H₄)_x·xLi.

The issue of mechanical property changes over time has been addressed and a structural model has been developed (12). A correlation was established between stiffness and the size of an endotherm (T_f), normally seen in dsc scans of ionomers at about 50°C. This endotherm increases in size with increasing neutralization. The T_f endotherm disappears completely when the dsc measurement is repeated immediately, but then gradually reappears during room-temperature storage. The heat of fusion in mJ /mg for this endotherm correlates with stiffness, as shown in Figure 2. It has been proposed that, at T_f , the ionic clusters undergo an order–disorder transition of the first order (12). The enhanced stiffness on neutralization is dependent on the formation of a rigid, ordered structure of an ionic salt group (12). It is also proposed that the ordered structure of ionic salts in ethylene ionomers varies depending on the types of metal ions, with three carboxyls around each cation for Group 1 (1A) metals and zinc, and six for magnesium, calcium, strontium, and barium. The suggested structures for sodium and magnesium crystallites are shown in Figure 3.

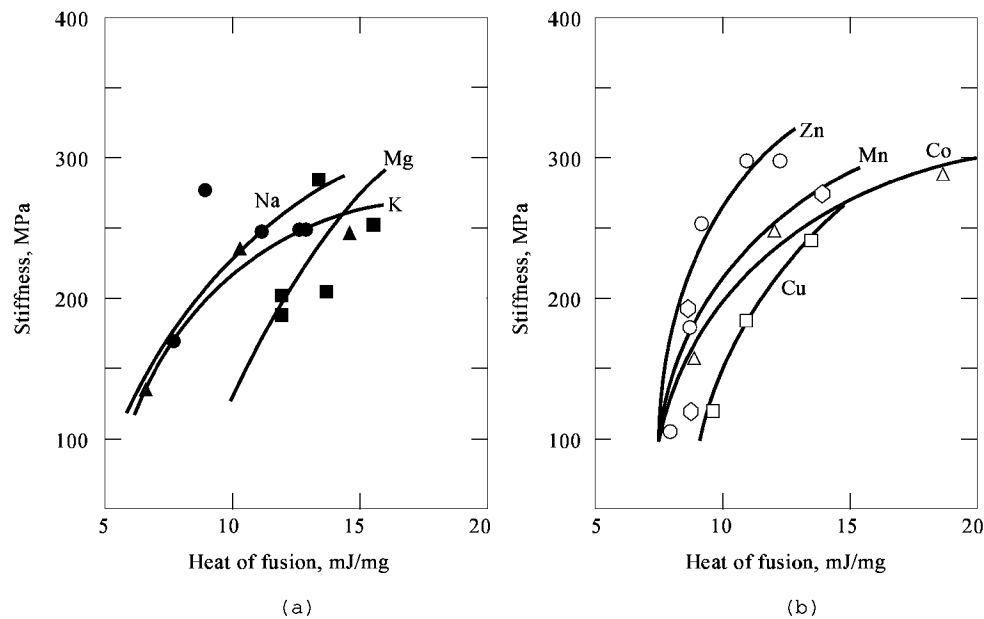


Fig. 2. Stiffness vs heat of fusion, ΔH_f , of ionic crystallites for ionomers. Cation type (a): (○) Na⁺, (▲) K⁺, (■) Mg²⁺; (b): (○) Zn²⁺, (△) Co²⁺, (□) Cu²⁺, (◇) Mn²⁺ (12). To convert mJ /mg to cal /g, divide by 4.184.

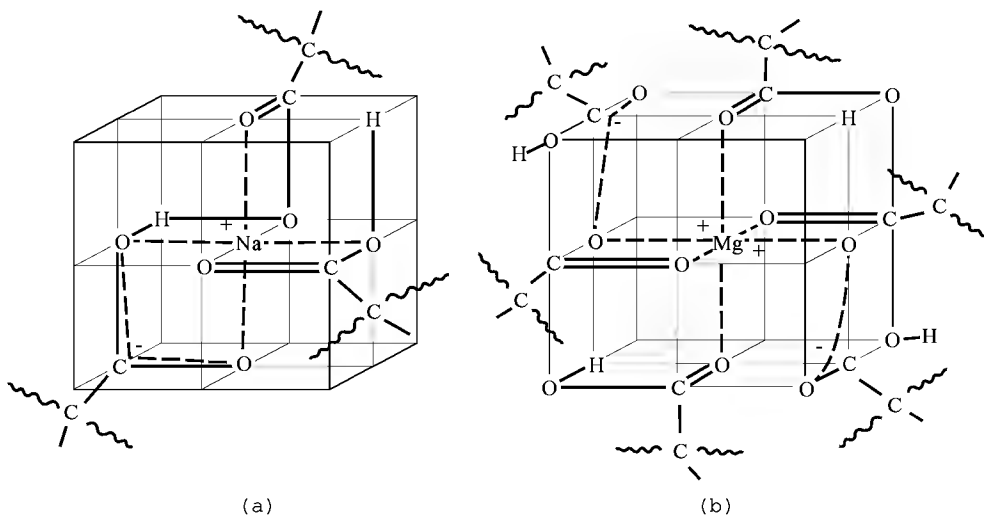


Fig. 3. Proposed structures of ionic crystallites for (a) sodium and (b) magnesium ionomers (12).

In addition to time-related effects, the solid-state physical properties are also affected by adsorbed water, which functions as a plasticizer. Water pickup is affected by the nature of the cation, with sodium ionomers absorbing about 10 times the level of the zinc equivalent (6) under the same conditions. Drying must be carried out at temperatures below 100°C and is therefore a slow process. In commercial practice, ionomers are supplied dry, and techniques have been developed to minimize moisture absorption during processing.

Crystallinity of Ionomers. Ionomers are much less hazy than the ethylene acid copolymers from which they are derived (6,11). Studies with optical and electron microscopes have shown that this is due to suppression of the spherulitic structure by the metal ions (11,14). Surprisingly, x-ray diffraction has shown that polyethylene crystallinity is present in the ionomers (11,15). X-ray diffraction measurements over a range of temperatures give a melting point similar to that obtained by dsc, a substantial 5°C lower than that of the acid copolymer. Calorimetric studies indicate that supercooling is more important in the ionomers than in acid copolymers (16,17). The x-ray work indicates that the degree of crystalline perfection in the ionomers is slight, and the strong ionic forces prevent the formation of comparatively large lamellae and spherulites, although the basic folded chain segments are present. A typical level of crystallinity is 30% (6).

Rheological Properties.

Melt Viscosity. As shown in Tables 2 and 3, the melt viscosity of an acid copolymer increases dramatically as the fraction of neutralization is increased. The relationship for sodium ionomers is shown in Figure 4 (6). Melt viscosities for a series of sodium ionomers derived from an ethylene–3.5 mol % methacrylic acid polymer show that the increase is most pronounced at low shear rates and that the ionomers become increasingly non-Newtonian with increasing neutralization (9). The activation energy for viscous flow has been reported to be somewhat higher in ionomers than in related acidic

copolymers (6,9,11).

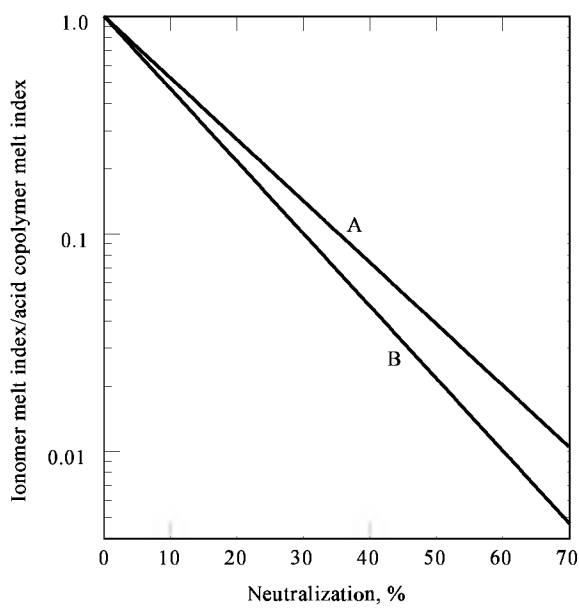


Fig. 4. Ionomer melt index reduction vs percent neutralization for sodium ionomers. A, 10 wt % methacrylic acid; B, 15 wt % methacrylic acid (6).

Since the melt viscosities of ionomers can be controlled by adjusting the polymer molecular weight and the level of neutralization, commercial products having melt-flow values similar to those of conventional polyolefins are available. However, the melt elasticity and melt strength characteristics of ionomers are very different from those of polyethylene resins. In high speed extrusion operations involving drawing of the polymer in the molten state, the resistance of ionomers to tear-off is unusually high. This has been attributed to low segmented mobility in the ionic clusters and higher energy storage (9). The principal effect of neutralization is to create relaxation processes at longer times. The retardation time goes from 100 to 10⁵ s, owing to an increase in internal viscosity from 10⁵ to 10⁸ Pa s (10⁵ to 10⁹ P).

Very high melt strength is also observed in situations where the molten polymer contacts a sharp object. Skin packaging trials have successfully packaged objects such as arrowheads and fishhooks in ionomer films.

Hot tack strength is the ability of a heat-seal layer to hold together while molten, before the seal cools and sets up. This is a technically important property which is difficult to measure reproducibly in the laboratory. Owing to the reinforcing effect of ionic bonding on melt strength, ionomer sealing layers provide superior performance in a wide spectrum of applications.

Melting Behavior. Ethylene acid copolymers resemble low density polyethylenes in their macroscopic behavior as the temperature is raised, with a relatively sharp transition from a flexible solid to a liquid melt. The transition is much more gradual in the ionomers, with softening apparent over a wide range, while the melt is strong and elastic. In low stress heat-distortion tests, molded ionomer samples begin distorting at lower temperatures than acid copolymers as the temperature increases, but retain their shapes at much higher temperatures (11). This gradual melting is beneficial in heat-sealing applications. Strong seals can be obtained in commercial applications involving variations in sealing-bar temperatures and contamination.

A dsc scan of a typical commercial ionomer shows two endotherms at about 50 and 98°C, respectively. The size of the lower peak can be correlated with stiffness and yield point. The thermal history of the sample influences the relative size of the lower peak and moves it to higher temperatures, while the upper peak decreases in size but remains at the same temperature. Room temperature aging also increases the size of the lower endotherm.

X-ray diffraction work (11,15) shows that there is an ionomer peak at 4°C which is absent in the acid precursor. This low, broad peak is not affected by annealing or ion type and persists up to 300°C. Since the 4°C peak corresponds to a spacing of about 2.5 nm, it is reasonable to propose a structural feature of this dimension in the ionomer. The concept of ionic clusters was initially suggested to explain the large effects on properties of relatively sparse ionic species (1). The exact size of the clusters has been the subject of much debate and has been discussed in a substantial body of literature (3,4,18–20). A theoretical treatment has shown that various models can give rise to supramolecular structures containing ionic multiplets which are about 10 nm in diameter (19).

Infrared Spectra of Ionomers. Infrared absorption data, first published in 1964 (11), show that partial neutralization of ethylene–methacrylic acid introduced new absorption bands at 1480–1670 cm^{–1} for the ionized carboxylate group while the 1698 cm^{–1} band of the free acid carboxyl diminishes in size (21). In addition to providing information on structural features, the numerous absorption bands are significant in applications technology, providing rapid warmup of film and sheet under infrared radiation.

Although all ionomers exhibit absorption bands in the general area of 1500–1620 cm^{–1}, the exact locations, shapes, and sizes differ according to the type of metal ion. It is possible to differentiate a zinc from a sodium ionomer by inspection of the infrared spectra since the zinc has a narrow peak, at 1585 cm^{–1}, and the sodium has a broad band, at 1520–1620 cm^{–1}. One study (21) of a wide variety of ionomer types, all based on a single starting copolymer, has resulted in two formation mechanisms and models related to these infrared peak differences. Zinc and copper, which both give sharp single peaks, are postulated to form coordination complexes, which are stable symmetrical structures. The coordination number for zinc is given as 4. However, in the alkali and alkaline-earth elements, there are both clusters and multiplets each having a distinct infrared peak, with a dynamic thermal equilibrium between them. The multiplet peak is at a lower wave number than that of the cluster (1547 vs 1568 cm^{–1} for Na⁺) and typically becomes larger relative to the room temperature multiplet. When the spectrum is recorded at 70°C, this study indicates a basic difference in structure between sodium and zinc ionomers, which together are by far the most important commercial types. Despite the structural differences, these ionomer types are similar in most melt and solid-state properties.

Solubility of Ionomers. Ionic bonding with metal ions decreases solubility in organic solvents (6,11). Commercial ionomers can generally be swollen by certain solvents such as aromatic hydrocarbons at elevated temperatures, but do not dissolve completely to give viscous solutions. Resistance to surface etching by organic solvents is high in most cases.

At high neutralization levels with alkali metal ions, many ionomers spontaneously form colloidal suspensions in water when stirred vigorously at 100–150°C under pressure. Depending on solids content and acid level, the dispersions range in viscosity from water-like to paste-like. These provide convenient methods for applying thin coatings of ionomers to paper and other substrates.

Electrical Properties. Due to the comparatively low content of polar groups, most commercial ionomers are very good insulating resins. Typical electrical properties (6) for a zinc ionomer are as follows:

Property	Value
dielectric constant	2.5
dissipation factor	0.0018
volume resistivity, Ω cm	0.53 × 10 ¹⁷

dielectric strength, V μ m 40.2

When exposed to high voltage electrical fields, many ionomers exhibit better resistance to the growth of dendrites (trees) than polyethylene. Due to the presence of ionically charged species, charge transfer by surface contact is facilitated, so high static charges can be generated when sheets or tubes are passed over rolls. This behavior can be a problem or an advantage, depending on the end use application.

Permeability. Ionic bonding has an important influence on permeability characteristics, especially where oily materials are involved. Acid copolymers are less permeable to natural oils than conventional homopolymers, and this difference increases greatly when they are neutralized, as illustrated in Table 4 (6).

In the area of gas permeability, the low crystallinity of a typical ionomer ($\sim$ 30%) results in relatively high permeability to oxygen. For packaging of fresh meat this is advantageous, but in other packaging areas, combination with a barrier layer may be required (see BARRIER POLYMERS).

Table 4. Oil Resistance of Ionomer Resin

Sample	Cottonseed oil, ^a h to failure	Peanut oil, ^a h to failure
high density polyethylene	40	25
ethylene–methacrylic acid	64	78
55 ^o neutralized copolymer	229	182

^a Oil breakthrough: 25 μ m coating at 60°C.

MANUFACTURE AND PROCESSING

Most commercial processes involve copolymerization of ethylene with the acid comonomer followed by partial neutralization, using appropriate metal compounds. The copolymerization step is best carried out in a well-stirred autoclave with continuous feeds of all ingredients and the free-radical initiator, under substantially constant environment conditions (22–24). Owing to the relatively high reactivity of the acid comonomer, it is desirable to provide rapid end-over-end mixing, and the comonomer content of the feed is much lower than that of the copolymer product. Temperatures of 150–280°C and pressures well in excess of 100 MPa (1000 atm) are maintained. Modifications on the basic process described above have been described (25,26). When specific properties such as increased stiffness are required, nonrandom copolymers may be preferred. An additional comonomer, however, may be introduced to decrease crystallinity (10,27).

Many methods for the conversion of acid copolymers to ionomers have been described by Du Pont (27,28). The chemistry involved is simple when cations such as sodium or potassium are involved, but conditions must be controlled to obtain uniform products. Solutions of sodium hydroxide or methoxide can be fed to the acid copolymer melt, using a high shear device such as a two-roll mill to achieve uniformity. All volatile by-products are easily removed during the conversion, which is run at about 150°C. A continuous process has been described, using two extruders, the first designed to plasticate the feed polymer and mix it rapidly with the metal compound, eg, zinc oxide, at 160°C (28). Acetic acid is pumped into the melt to function as an activator. Volatiles are removed in an extraction-extruder which follows the reactor-extruder, and the anhydrous melt emerges through a die-plate as strands which are cut into pellets.

An unusual slurry process which works well with sodium hydroxide is based on diffusion of the aqueous reagent into pellets of acid polymer (28). The concentration of ions in the liquid phase is preferably two to four times the stoichiometric level, and the temperature is maintained at 50–100°C.

A process based on saponification of ethylene–acrylate ester copolymers has been practiced commercially in Japan (29). The saponification naturally produces fully neutralized polymer, and it is then necessary to acidify in order to obtain a partly neutralized, melt-processible product. Technology is described to convert the sodium ionomer produced by this process to the zinc type by soaking pellets in zinc acetate solution, followed by drying (29).

For small-scale preparation of samples for scientific studies, the precursor polymer may be dissolved in xylene at 80°C, followed by addition of the cation source. A gelled fluid is normally obtained immediately, and the ionomer is recovered as a powder by chopping the gel in a large excess of acetone using a laboratory blender.

A direct method for obtaining a sodium ionomer by polymerizing a mixture of ethylene, sodium methacrylate, and methacrylic acid has been described (30).

Economic Aspects. Woddwide production is of the order of 110,000 t. The selling price ranges from U.S. \$2.8–4.8 kg, depending on grade and end use.

Specifications, Tests, and Shipment

Ionomer resins are produced in multiple grades to meet market needs, and prospective customers are provided with information on key processing parameters such as melt-flow index. Nominal values for many other properties are listed in product brochures. The ASTM test methods developed for general-purpose thermoplastic resins are applicable to ionomers. No special methods have been introduced specifically for the ionomers.

Ionomers are generally shipped in pellet form in the standard containers developed for large-volume polyolefins, eg, 500-kg boxes. Water-resistant liners are used to keep the products dry during shipment and storage.

Health and Safety Factors

During processing at elevated temperatures, normal precautions are needed to prevent accidental burns. Surlyn ionomers have U.S. Food and Drug Administration clearance for food contact. Information about ionomers can be found in the articles ETHYLENE; ACRYLIC ACID AND DERIVATIVES; and METHACRYLIC ACID AND DERIVATIVES.

Uses

Flexible packaging is the largest commercial application area for ethylene ionomers. Uses include monolayer films, coextruded components of multilayer films, and coatings on paperboard or aluminum. The key properties are broad heat-sealing range, ability to seal through oily contamination, high melt strength, oil resistance, clarity, and impact resistance.

For skin packaging, ionomer films provide outstanding protection, especially for sharp objects. Films can be drawn down tightly around regular shapes without penetration. Characteristics of heat seals can be adjusted so that convenient peelable seals are obtained (31–35).

The unusual resilience of ionomers combined with ease of processing have resulted in widespread replacement of balata rubber as golf-ball covers. In order to obtain desirable backspin characteristics, low glass-transition ionomer compositions have been developed (10). Transparent coatings on bowling pins promote both longer life and improved playing performance.

Low temperature formability and resilience are key properties in the area of shoe components, ie, box toes and counters. These ionomer components are not seen by the purchaser of the finished article, but provide comfort and durability in high quality footwear.

The transparency, solvent resistance, and attractive feel of ionomer moldings have resulted in a substantial European market in stoppers for bottles containing expensive perfumes. This is a demanding application since no loss of perfume ingredients can be tolerated.

Ionomer tubing is used in nonelectric explosive ignition systems (36,37). Key characteristics are shock resistance and retention of static charge on the interior tube surface. The charge keeps a thin layer of explosive powder in place. Pipe linings make use of improved abrasion resistance for pumping of abrasive slurries (38).

Transparent ionomer coatings are applied to glass surfaces to improve safety characteristics (39). Special types of ionomers are used as interlayers between glass sheets for applications requiring penetration resistance (40).

Ionomers are easily foamed due to high melt strength and the foams are durable, leading to uses in construction, skilifts, and softball cores (41). Blends of ionomers with other polymers are used commercially for molding parts that exhibit toughness. In many cases, the primary advantage to be obtained is increased impact resistance, but other functions such as controlling the crystallization rate are also important (42–44).

Noncommercial Ethylene-Based Ionomers

Amine-Linked and Complexed Ionomers. Organic bases, notably diamines, can be substituted for metal ions to give ionomers which have similar solid-state properties to those neutralized with metal ions but differ in the area of melt viscosity. A general overview of the various properties has been published (45,46). Diamines may also be combined with metal cations to give transparent, tough products (47–50). This technology is used commercially in glass interlayers.

Ethylene–Dicarboxylic Acid Copolymers. Partial neutralization of copolymers containing carboxyls in pairs on adjacent carbons, eg, ethylene–maleic acid, has been described (11). Surprisingly, there is no increase in stiffness related to neutralization. Salts with divalent metal cations are not melt processible. The close spacing of the paired carboxyl groups has resulted in ionic cluster morphology which is distinct from that of the commercial ionomer family.

Ionomers Not Based on Ethylene

Styrene-Based Ionomers. There are two main types of styrene-based ionomers, those derived from sulfonated polystyrene, or from copolymers of styrene with methacrylic acid, respectively (51). Both have been the subject of much academic study, although no commercial interest in either has developed. The properties of styrene-*co*-methacrylic acid ionomers have been presented (51–54). The results have been interpreted in terms of simple multiplets of ions at low concentrations, with larger clusters present at >6 mol% (51). Studies of sulfonated styrene ionomers show that the strength of ionic association is much higher than that of the carboxylated polymers, and the melt viscosity is extremely high (55–57). The polymers can be plasticized by glycerol or other additives (55). Styrene-based ionomers having ionic groups separated from the backbone have also been prepared and evaluated (58).

EPDM-Derived Ionomers. Another type of ionomer containing sulfonate, as opposed to carboxyl anions, has been obtained by sulfonating ethylene–propylene–diene (EPDM) rubbers (59,60). Due to the strength of the cross-link, these polymers are not inherently melt-processible, but the addition of other metal salts such as zinc stearate introduces thermoplastic behavior (61,62). These interesting polymers are classified as thermoplastic elastomers (see ELASTOMERS, SYNTHETIC THERMOPLASTIC ELASTOMERS).

Butadiene–Methacrylic Acid Ionomers. Carboxyl groups can readily be introduced into butadiene elastomers by copolymerization, and the effects of partial neutralization have been reported (63–66). The ionized polymers exhibit some degree of fluidity at elevated temperatures, but are not thermoplastic elastomers, and are very deficient in key elastomer properties such as compression set resistance.

Telechelic Ionomers. Low molecular weight polymers terminated by acid groups have been treated with metal bases to give ionomers in which the cations can be considered as connecting links in the backbones (67–71). The viscoelastic behavior of concentrated solutions has been linked to the neutralizing cation.

Pentenamer Ionomers. Unsaturated polypentenamer elastomers have been derivatized by post-synthesis reactions (72–74). Phosphonate, thioglycolate, sulfonate, and carboxylate derivatives have been prepared and converted into ionomers.

Bitumen Ionomers. Moisture-resistant asphalts (qv) have been prepared by reaction of metal oxides with acid-functionalized bitumens (75). Maleic anhydride or sulfur trioxide trimethylamine complexes have been used successfully for introduction of acid groups into asphaltic bitumens.

Polyoxymethylene Ionomers. Ionic copolymers have been prepared from trioxane and epichlorohydrin, followed by reaction with disodium thioglycolate (76). The ionic forces in these materials disrupt crystalline order and increase melt viscosity (see ACETAL RESINS).

New Ionomer Types. There is a continuing interest in new ionomers within the academic community, since novel and unexpected phenomena are frequently being discovered. However, there are still many unanswered questions with respect to the ethylene ionomers, especially the influence of ionic bonding on crystalline structure. Continued study of these interesting polymers will close the gaps in knowledge of this area of polymer science.

BIBLIOGRAPHY

"Ionomers" in *ECT* 3rd ed., Suppl. Vol., pp. 546–573, by R. D. Lundberg, Exxon Research and Engineering Co.

1. R. W. Rees, *Mod. Plast.* **42**, 209 (Sept. 1964).
2. D. C. Watts, *Chem. Brit.* **28**, 238 (Mar. 1992).
3. A. Eisenberg, ed., *Ions in Polymers*, ACS Monograph No. 187, American Chemical Society, Washington, D.C., 1980.
4. W. J. MacKnight and T. R. Earnest, *J. Macromol. Rev.* **16**, 41 (1981).
5. L. Holliday, ed., *Ionic Polymers*, Halstead Press, a division of John Wiley & Sons, Inc., New York, 1975.
6. R. W. Rees, in K. C. Frisch, ed., *Polyelectrolytes*, Technomic Publishing Co., Inc., Westport, Conn., 1976, pp. 177–197.
7. N. L. Zutty, J. A. Faucher, and S. Bonotto, in N. M. Bikales, ed., *Encyclopedia of Polymer Science and Technology*, Vol. 6, Interscience Publishers, a Division of John Wiley & Sons, Inc., New York, 1967, p. 420.
8. R. W. Rees, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, 2nd ed., Vol. 4, Wiley-Interscience, New York, 1986, pp. 395–417.
9. R. Longworth, *Plast. Rubber*, 75 (Aug. 1978).
10. **U.S. Pat. 4,690,981** (Sept. 1, 1987), R. J. Statz (to E. I. du Pont de Nemours & Co., Inc.).
11. R. W. Rees and D. J. Vaughan, *Polym. Prep. Amer. Chem. Soc. Div. Polym. Chem.* **6**, 287–303 (1965).
12. E. Hirasawa, Y. Hamamoto, K. Tadano, and S. Yano, *J. Appl. Polym. Sci.* **42**, 351 (1991).
13. S. Bonotto and E. F. Bonner, *Macromolecules* **1**, 510 (1968).
14. H. A. Davis, R. Longworth, and D. J. Vaughan, *Polym. Prep. Amer. Chem. Soc. Div. Polym. Chem.* **9**, 515 (1968).
15. F. C. Wilson, R. Longworth, and D. J. Vaughan, *Polym. Prep. Amer. Chem. Soc. Div. Polym. Chem.* **9**, 505 (1968).
16. T. Kajiyama, T. Oda, R. S. Stein, and W. J. MacKnight, *Macromolecules* **4**, 198 (1971).
17. C. L. Marx and S. L. Cooper, *Makromol. Chem.* **168**, 339 (1973).
18. C. G. Bazuin and A. Eisenberg, *Ind. Eng. Chem.* **20**, 271 (1981).
19. A. Eisenberg, *Macromolecules* **3**, 147 (1970).
20. A. Neppel, I. S. Butler, N. Brockman, and A. Eisenberg, *J. Macromol. Sci.* **B19**, 61 (1980).
21. K. Han and H. L. Williams, *J. Appl. Polym. Sci.* **42**, 1845 (1991).
22. **U.S. Pat. 4,351,931** (Sept. 28, 1982), J. B. Armitage (to Du Pont).
23. **U.S. Pat. 3,520,861** (July 21, 1970), J. E. Thomson and G. C. Waples (to the Dow Chemical Company).
24. **U.S. Pat. 4,599,392** (July 8, 1986), O. K. McKinney, D. P. Flores, and D. A. Eversdyk (to the Dow Chemical Company).
25. **U.S. Pat. 4,252,924** (Feb. 24, 1981), A. Chatterjee (to Du Pont).

26. U.S. Pat. **4,248,990** (Feb. 3, 1981), E. T. Pieski and T. F. Sashihara (to Du Pont).
27. U.S. Pat. **3,264,272** (Aug. 2, 1966), R. W. Rees (to Du Pont).
28. U.S. Pat. **3,404,134** (Oct. 1, 1968), R. W. Rees (to Du Pont).
29. U.S. Pat. **3,789,035** (Jan. 29, 1974), J. Iwami, H. Kawasaki, H. Kinoshita (to Asahi-Dow Ltd.).
30. Eur. Pat. Appl. **223,182** (May 27, 1987), J. Herwig, A. Kolwest, H. Sutter, D. Woltjes (to E. D. Erdoelchemie GmbH).
31. U.S. Pat. **4,188,441** (Feb. 12, 1980), S. G. Cook (to Crown Zellerbach).
32. U.S. Pat. **4,346,196** (Aug. 24, 1892), G. L. Hoh and D. A. Vassallo (to Du Pont).
33. U.S. Pat. **4,469,754** (Sept. 4, 1984), G. L. Hoh and D. A. Vassallo (to Du Pont).
34. U.S. Pat. **4,539,263** (Sept. 3, 1985), G. L. Hoh (to Du Pont).
35. U.S. Pat. **4,550,141** (Oct. 29, 1985), G. L. Hoh (to Du Pont).
36. U.S. Pat. **3,590,739** (July 6, 1971), P. A. Perssor (to Nitro Nobel).
37. U.S. Pat. **3,125,024** (Mar. 17, 1964), M. A. Hicks and W. Mann (to Can Safety Fuse, Inc.).
38. R. J. Statz, *Polym. Prep. Amer. Chem. Soc. Div. Polym. Chem.* **29**, 435 (1988).
39. U.S. Pat. **4,056,208** (Nov. 1, 1977), G. W. Prejean (to Du Pont).
40. U.S. Pat. **4,619,973** (Oct. 28, 1986), N. W. Smith (to Advanced Glass Systems).
41. U.S. Pat. **4,572,507** (Feb. 25, 1986), D. W. Hubbert and B. L. Robinson (to Athlone Industries).
42. U.S. Pat. **4,174,358** (Nov. 13, 1979), B. N. Epstein (to Du Pont).
43. U.S. Pat. **4,346,194** (Aug. 24, 1982), M. J. Roura (to Du Pont).
44. U.S. Pat. **4,352, 904** (Oct. 5, 1982), E. J. Deyrup (to Du Pont).
45. R. W. Rees, *Polym. Prep. Amer. Chem. Soc. Div. Polym. Chem.* **14**, 796 (1973).
46. U.S. Pat. **3,471,460** (Oct. 7, 1969), R. W. Rees (to Du Pont).
47. E. Hirasawa, Y. Yamamoto, K. Tadano, and S. Yano, *Macromolecules* **22**, 2776 (1989).
48. S. Yano, H. Yamamoto, K. Tadano, Y. Yamamoto, and E. Hirasawa, *Polymer* **28**, 1965 (1987).
49. K. Tadano, E. Hirasawa, H. Yamamoto, and S. Yano, *Macromolecules* **22**, 226 (1989).
50. U.S. Pat. **4,732,944** (Mar. 22, 1988), W. N. Smith (to Advanced Glass Systems).
51. A. Eisenberg and M. King, *Ion-Containing Polymers*, Academic Press, Inc., New York, 1977.
52. A. Eisenberg and M. Navratil, *Macromolecules* **6**, 604 (1973).
53. *Ibid.*, **7**, 90 (1974).
54. U.S. Pat. **3,322,734** (May 30, 1967), R. W. Rees (to Du Pont).
55. R. D. Lundberg, H. S. Makowski, and L. Westerman, *Adv. Chem. Ser.* **187**, 67 (1980).
56. R. D. Lundberg and H. S. Makowski, *Adv. Chem. Ser.* **187**, 21 (1980).
57. U.S. Pat. **3,870,841** (1975), H. S. Makowski, R. D. Lundberg, and G. S. Singhal (to Exxon).
58. R. B. Moore, D. Bittencourt, M. Gauthier, C. Williams, and A. Eisenberg, *Macromolecules* **24**, 1376 (1991).
59. H. S. Makowski and R. D. Lundberg, *Adv. Chem. Ser.* **187**, 37 (1980).
60. H. S. Makowski, R. J. Lundberg, L. Westerman, and J. Bock, *Adv. Chem. Ser.* **187**, 3 (1980).
61. U.S. Pat. **3,847,854** (Nov. 12, 1974), N. H. Canter and D. J. Buckley (to ESSO).
62. R. A. Weiss, *J. Appl. Polym. Sci.* **28**, 3321 (1983).
63. H. P. Brown, *Rubber Chem. Technol.* **30**, 1347 (1957).
64. H. P. Brown and C. F. Gibbs, *Rubber Chem. Technol.* **28**, 937 (1955).
65. W. Cooper and T. B. Bird, *Ind. Eng. Chem.* **50**, 771 (1958).
66. W. Cooper, *J. Polym. Sci.* **28**, 195 (1958).
67. G. Boze, R. Jerome, and P. Teyssie, *Macromolecules* **14**, 224 (1981).
68. G. Boze, R. Jerome, P. Teyssie, and C. Marco, *Polym. Bull. Berlin* **4**, 241 (1981).
69. G. Boze, R. Jerome, P. Teyssie, and B. Gallot, *J. Polym. Sci. Polym. Lett. Ed.* **19**, 415 (1981).
70. G. Boze, R. Jerome, and P. Teyssie, *Macromolecules* **15**, 920 (1982).
71. G. Boze, R. Jerome, P. Teyssie, and C. Marco, *Macromolecules* **16**, 996 (1983).
72. C. Azuma and W. J. MacKnight, *J. Polym. Sci. Polym. Chem. Ed.* **15**, 547 (1977).
73. K. Sanui, R. W. Lenz, and W. J. MacKnight, *J. Polym. Sci. Polym. Chem. Ed.* **12**, 1965 (1974).
74. D. Rahrig and W. J. MacKnight, *Polym. Prep. Amer. Chem. Soc. Div. Polym. Chem.* **19**, 314 (1978).
75. L. Ciplijauskas, M. R. Piggott, and R. T. Woodhams, *Adv. Chem. Ser.* **187**, 171 (1980).
76. K. F. Wissbrun, *Makromol. Chem.* **118**, 2112 (1968).

Richard W. Rees
E. I. du Pont de Nemours & Co., Inc.

IONOPHORES.

See ANTIBIOTICS, POLYETHERS; ANTIBIOTICS, PEPTIDES; CHELATING AGENTS.

ION-SELECTIVE ELECTRODES.

See ELECTROANALYTICAL TECHNIQUES.

IRIDIUM.

See PLATINUM GROUP METALS.

IRON

Iron [7439-89-6], Fe, from the Latin *ferrum*, atomic number 26, is the fourth most abundant element in the earth's crust, outranked only by aluminum, silicon, and oxygen. It is the world's least expensive and most useful metal. Although gold, silver, copper, brass, and bronze were in common use before iron, it was not until humans discovered how to extract iron from its ores that civilization developed rapidly (see MINERAL PROCESSING AND RECOVERY).

Pure iron is a silvery white, relatively soft metal and is rarely used commercially. Typical properties are listed in Table 1. Electrolytic (99.9% pure) iron is used for magnetic cores (2) (see MAGNETIC MATERIALS, BULK). Native metallic iron is rarely found in nature because iron which commonly exhibits valences of +2 and +3 combines readily with oxygen and sulfur. Iron oxides are the most prevalent form of iron (see IRON COMPOUNDS). Generally, these iron oxides (iron ores) are reduced to iron and melted in a blast furnace. The hot metal (pig iron) from the blast furnace is refined in steelmaking furnaces to make steel (qv).

Iron is alloyed with other elements for commercial applications. The most important alloying element is carbon. Small amounts of carbon alloyed with iron lower the melting point, as illustrated in Figure 1. The distinction between steels and other irons is based on properties and defined by the iron–carbon phase diagram. Steel is generally classified as those iron–carbon alloys (0–2% C) which have a high melting point and can be hot rolled. Iron with carbon up to about 2% can be heated to a temperature at which only one phase (gamma iron) exists. Gamma iron is face-centered cubic (fcc) in structure, and therefore is plastic, or malleable, which allows hot rolling. Cast irons are those which contain sufficient quantities of the eutectic (about 2–5% C) to make the metal too brittle to hot roll; thus the requirement that it be cast. Pig iron from the blast furnace is liquid iron saturated with carbon (>4.3% C) depending on the temperature corresponding to the liquidus line.

Table 1. Properties of Iron^a

Property	Value	
atomic mass	55.847	
isotopic abundance		
mass	54	56 57 58
abundance, %	6.04 91.57 2.11 0.28	
melting point, °C	1537	
boiling point, °C	3000	
crystal structure ^b	bcc	
density, g cm ³	7.87	
thermal conductivity at 0°C, W (mK)	79	
electrical resistivity at 20°C, μΩcm	9.71	
tensile strength, MPa ^d	240–280	
yield strength, MPa ^d	70–140	
Young's modulus of elasticity, GPa ^d	195	
Poisson's ratio	0.3	
elongation in 5 cm at 20°C, %	40–60	
reduction of area, %	65–78	
Brinell hardness	82–100	
impact strength (izod notched bar)		
longitudinal, J m ^e	4859	
transverse, J m ^e	2990	
thermal expansion, K ⁻¹		
from 0–300°C	12.6 × 10 ⁻⁶	
0–600°C	14.6 × 10 ⁻⁶	
specific heat, J (gK) ^g		
at 100°C	0.50	
500°C	0.67	
800°C	1.26	
transition from magnetic to paramagnetic, °C	ca 770	

^a Ref. 1.
^b Room temperature.
^c Hot rolled.
^d To convert MPa to psi, multiply by 145.
^e To convert J to cal, divide by 4.184.

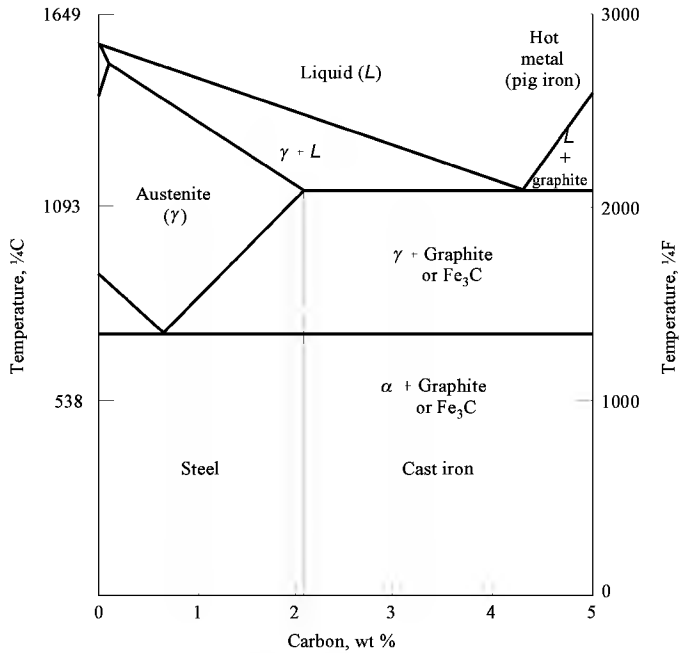


Fig. 1. Iron–carbon phase diagram, where α is the body-centered cubic (bcc) α -iron, γ is the face-centered cubic γ -iron, and Fe_3C is iron carbide(3:1) [12011-67-5] at 6.69° C. The vertical dashed line represents the demarcation between steel on the left and cast iron on the right (3).

Iron is indispensable in the human body (see MINERAL NUTRIENTS). The average adult body contains 3 grams of iron. About 65% is found in hemoglobin, which carries oxygen from the lungs to the various parts of the body. Iron is also needed for the proper functioning of cells, muscles, and other tissues (4).

The earliest record of human usage of iron dates to ca 2000 BC (5) in Egypt, Asia Minor, Assyria, China, and India. It is almost certain, however, that the first iron to be used was not processed but was obtained from meteorites (1). One of the few places where native iron is found is in Greenland, where it occurs as very small grains or nodules in basalt (an iron-bearing igneous rock) that erupted through beds of coal.

Processed iron was first produced around 1300 BC. It is presumed that the first iron was made accidentally as a result of very hot fires built on top of some iron-bearing rocks or soil. The iron-bearing rocks could have been reduced to iron by being heated in the presence of hot charcoal and in the absence of air. Upon raking out the ashes, the first ironmaker probably found a sponge-like chunk of hard but malleable metal having considerable slag in its pores. Reduced iron sponge had to be hammered and squeezed while still hot to expel most of the slag in order to make effective use of the metal. This hammering and working process produced wrought iron.

The first furnaces specifically made for smelting iron ore were low shaft furnaces: low box-like hearths made of stone, open at the top, and having an opening near the bottom for air intake. The Catalan hearth furnace for making wrought iron is described in writings from Central Europe in the twelfth and thirteenth centuries AD. These Catalan furnaces were low stone shaft furnaces having hearth dimensions of ca 60 × 60 × 75 cm. As better blowing devices were invented and the height of the furnaces was increased, it is probable that, when the fires became hot, liquid high carbon iron, which was not malleable, was produced and could be used to make cast-iron articles.

The Stuckofen or old high bloomery appeared in Germany in ca 1300 AD. This type of furnace was 3–5 m high and enclosed a tapered vertical shaft that was 1–1.2 m in diameter. Small openings near the bottom were provided for nozzles (tuyeres, pronounced *tuvers*) that permitted air, supplied by bellows, to be blown into the furnace. Modern blast furnaces have essentially the same fundamental design.

Eventually, processes were developed for converting the high carbon product of the blast furnace to wrought iron. In the puddling process, solidified pig iron first was melted and then silicon, manganese, and carbon were oxidized by the hot gases and the iron oxides that were added. During this operation the puddler continually stirred the bath. As the decarburization approached completion, the metal began to solidify. A series of hammering and squeezing operations separated the slag from the iron to produce wrought iron.

In the United States, the first ironworks was built at Jamestown, Virginia, in 1619. The Hammersmith furnace in Saugus, Massachusetts, built in 1645, operated until 1675. This early American ironworks has been restored and is called the Saugus Iron Works. Iron blast furnaces appeared in many localities where there were deposits of iron ore. Small bodies of iron ore in New Jersey, Connecticut, Massachusetts, Pennsylvania, and New York formed the basis of many small colonial furnaces.

Ironmaking in the United States did not expand rapidly until after the Revolutionary War. Then, as the colonists moved westward, the need for iron prompted the establishment of ironworks near the new settlements. A blast furnace built by Jacob Anschutz in 1796 was the beginning of the iron and steel center in Pittsburgh, Pennsylvania.

The early U.S. blast furnaces of the nineteenth century were in the form of a truncated cone or pyramid, 6–9 m in height, enclosing a shaft 1.2–2.4 m in maximum diameter, and constructed of stone. The output of these furnaces was from 1–6 metric tons per day. Many changes were made in furnace design and operation: the shaft was enclosed in a metal shell, the top was closed to prevent the escape of top gases, the blast air was preheated, and the furnaces were enlarged. By 1900, a single blast furnace made as much iron as 200 t d. Improvements continued to be made, but generally modern blast furnaces are only larger than their nineteenth century ancestors. The modern blast furnace may have a hearth as much as 14 m in diameter, be over 60 m in height, and produce up to 10,000 t d.

In 1979, there were 168 blast furnaces in the United States, most located in Pittsburgh and Chicago, and these produced ca 8×10^7 t of pig iron. By 1992, most of the blast furnaces in the Pittsburgh area had disappeared. Only 44 blast furnaces were operating in the United States, producing ca 4.7×10^7 t of pig iron. The drop in pig iron production can be attributed to decreased and more efficient use of steel products, competition from steel imports, and rapid growth of scrap-based steelmaking.

Iron Ores

Minerals. Iron-bearing minerals are numerous and are present in most soils and rocks. However only a few minerals are important sources of iron and thus called ores. Table 2 shows the principle iron-bearing minerals. Hematite is the most plentiful iron mineral mined, followed by magnetite, goethite, siderite, ilmenite, and pyrite. Siderite is unimportant in the United States, but is an important source of iron in Europe. Ilmenite is normally mined for titania with iron as a by-product. Pyrite is roasted to recover sulfur in the form of sulfur dioxide, leaving iron oxide as a by-product.

Table 2. Principal Iron-Bearing Minerals^a

Parameter	Mineral					
	Hematite	Magnetite	Goethite	Siderite	Ilmenite	Pyrite
CAS Registry Number	[1309-37-1]	[1309-38-2]	[1310-14-1]	[14476-16-5]	[12168-52-4]	[1309-36-0]
chemical name	ferric oxide	ferrous–ferric oxide	hydrous iron oxide	iron carbonate	iron–titanium oxide	iron sulfide
chemical formula	Fe ₂ O ₃	Fe ₃ O ₄	HFeO ₂	FeCO ₃	FeTiO ₃	FeS ₂
iron, wt %	69.94	72.36	62.85	48.20	36.80	46.55
color	steel gray to red	dark gray to black	yellow or brown to nearly black	white to greenish gray to black	iron-black	pale brass-yellow
crystal	hexagonal	cubic	orthorhombic	hexagonal	hexagonal	cubic
specific gravity	5.24	5.18	3.3–4.3	3.83–3.88	4.72	4.95–5.10
Mohs' hardness	6.5	6	5–5.5	3.5–4	5–6	6–6.5
melting point, °C	1565	1600			1370	
magnetism		strong			slight	

^a Ref. 1.

Sources. Iron ore deposits were formed by many different processes, eg, weathering, sedimentation, hydrothermal, and chemical. Iron ores occur in igneous, metamorphic, and sedimentary deposits. Normally, as-mined iron ore contains 25 to 68% iron.

The main iron ore deposits in the United States lie near Lake Superior in Minnesota (Mesabi range) and Michigan (Marquette range). These deposits were formed by sedimentation during the Precambrian era. The original formation consisted of layers of iron oxides, iron carbonates, and iron silicates that were interbedded with layers of chert (a dense sedimentary rock). Leaching by groundwater and oxidation of the iron minerals produced local bodies of enriched oxides, usually soft hematites and limonites (goethites), that can be mined commercially without needing significant beneficiation. The unaltered iron formation, when containing 20% or more of iron in the form of magnetite, is called taconite; if in the form of hematite, it is referred to as jaspilite. Taconite and jaspilite are of importance because these are readily amenable to beneficiation. All of the mining in this area is done by open pit. The iron ore is beneficiated by magnetic separation and or flotation (qv) and pelletized (see SEPARATION, MAGNETIC).

Other iron ore deposits of lesser importance in the United States are located in Missouri, Utah, Alabama, Wyoming, Texas, California, Nevada, Pennsylvania, New York, New Jersey, and Wisconsin. Of these deposits, only Missouri and Utah have mines operating. The Missouri deposit is located southwest of St. Louis near Sullivan and is a steeply dipping igneous intrusion in surrounding rock. The ore is composed principally of magnetite with some hematite, and minor amounts of quartz, apatite, and pyrite. The crude ore is mined by underground methods and is upgraded from 56 to 70% iron by magnetic separation. The upgraded ore is used for specialty iron oxide applications such as pigments, ceramics (qv), and powdered metals.

The Utah deposit is located in southwestern Utah near Cedar City. The iron ore deposits are of contact metamorphic origin. The crude ore contains 35 to 65% iron, primarily in the form of magnetite and goethite. Mining is done by the open pit method. The crude ore is crushed, screened at 75 mm (200 mesh size) and shipped as lump ore containing 54% iron. The ore is rescreened at the steel mill to produce lump ore (10–64 mm) for the blast furnace and sinter feed (0–10 mm) for the sinter plant.

Canada's chief deposits occur along the borders between Quebec and Newfoundland in an area called the Labrador Trough, and in an area north of Lake Superior. Most of the deposits are similar to those found in Minnesota and Michigan. The Labrador Trough deposits are the main ones being mined. However, some direct shipping ore, containing mostly siderite, is produced in the Michipicoten range north of Lake Superior for use in sinter plants.

Other countries that have large iron ore deposits include Brazil (Carajas and Quadrilatero Ferrifero deposits), Australia (Pilbara deposits), Ukraine (Krivi Rog deposit), Russia (Kursk deposit), Venezuela (Cerro Bolivar deposit), India (Bihar-Orissa, Hospet, Kudremukh, and Goa deposits), South Africa (Sishen and Thabazimbi deposits), and Sweden (Kiruna, Svappavaara, and Malmberget deposits). A list of world iron ore production and reserves in 1992 by country is shown in Table 3.

Table 3. World Production and Reserves of Iron Ore in 1992, 10⁶ t^a

Country	Iron ore		Iron reserves
	Production	Reserves	
China	196	9,000	3,500
Brazil	146	17,300	10,100
Australia	117	28,100	17,900
Ukraine ^b	87	40,000	15,000
Russia ^b	83	38,000	14,000
United States	55	25,200	6,000
India	55	12,100	6,300
Canada	35	25,500	10,000
South Africa	28	9,300	5,900
Sweden	19	4,600	2,400
Venezuela	18	3,300	1,700
other countries	81	17,600	7,200
Total	920	230,000	100,000

^a Refs. 6 and 7.

^b Estimated value.

Annual world iron ore production has hovered around 9 to 9.75 × 10⁸ t since the mid-1980s. International trade of iron ore peaked in 1980 at 4.24 × 10⁸ t; otherwise it has remained fairly steady in the range of 3.6 to 4 × 10⁸ t yr per year since the mid-1980s. The main exporting countries are Brazil, Australia, India, Canada, South Africa, Russia, Ukraine, and Venezuela.

Beneficiation. Iron ore coming from the mine must be properly sized. A gyratory crusher is normally used for primary crushing down to approximately 300 mm (see SIZE REDUCTION). Secondary crushing down to 25 mm can be done in a cone crusher. Fine grinding can be done by rod mills followed by either ball or pebble mills. In some cases, autogenous grinding can be used to replace the cone crusher and rod mills.

High grade ore, containing over 60% total iron, can be used as direct shipping ore and normally is sized at 6–40 mm. High grade ore fines, in a size range of less than 6 mm, can be sold as sinter feed or further ground and agglomerated into pellets. Low grade iron ore must be ground to an acceptable size (to liberate gangue components) followed by concentration. The unit operations used for concentrating (or upgrading) low grade iron ores include washing, screening, heavy media separation, magnetic separation, froth flotation, and electrostatic separation (8).

Iron ores of different characteristics and compositions can be blended to a more uniform composition. This can be accomplished during handling operations involved in transporting ore to its point of use, or through special blending facilities, such as stacking and reclaiming.

Sand and clay can be removed from iron ore by washing in a log-washer or classifier followed by screening. If the sized ore consists of loose particles of mostly iron oxide mixed with loose particles of mostly gangue such as quartz and calcite, it is possible to upgrade the ore through gravity concentration based on differences in specific gravity. Heavy media separators using ferrosilicon suspension in a rotary drum are normally used for coarse-sized ore (10–25 mm). Heavy media separators using a magnetite suspension in hydraulic cyclones are used for medium-sized ore (0.8–6.7 mm). Humphrey’s spirals are used for upgrading fine ores (< 2.3 mm) (see SEPARATION, SIZE).

Low intensity magnetic separators are used to upgrade iron ores containing magnetite. Dry separators are used for coarse (up to 100 mm in size) ore and wet separators are used for fine (> 9.5 mm) ores.

High intensity magnetic separators are used to upgrade iron ores containing hematite or ilmenite. Dry separators require ore that is finely sized and bone dry. They are dusty, expensive, and have a low capacity. Wet separators have larger capacity, are less dusty and can handle ore sizes up to 1 mm.

Hematite or goethite can be converted to magnetite by reduction roasting at 500–550°C. Siderite can be converted to magnetite by roasting in a neutral atmosphere at 700–775°C. In both cases the roasted ore must be cooled in an air-free atmosphere to 100°C. The roasted ore is easier to crush and grind than unroasted ore and can be upgraded by low intensity magnetic separation. However, the cost of roasting usually is too expensive to justify.

Froth flotation can be used effectively to upgrade iron ore sized at less than 0.2 mm. Anionic flotation is used to float hematite or siderite away from quartz or chert. The reagent used is normally a fatty acid or petroleum sulfonate. Cationic flotation, employing amines as reagent, is used to float quartz and locked quartz–magnetite particles from clean magnetite particles. Electrostatic separation can be used to upgrade dry iron ores sized at 0.04–1.68 mm and free from slimes or dust coatings (see FLOTATION).

The flow sheet can be tailored to the specific characteristics of the ore. The unit operations can be combined or modified in many ways to upgrade the total iron content in the ore from levels as low as 20–38% up to levels of 65–70%.

Agglomeration. Iron ore concentrates are often too fine to be used directly in ironmaking processes; therefore they must be agglomerated. The agglomerating methods typically used in the iron ore industry are pelletizing, sintering and, to a limited extent, briquetting and nodulizing (see SIZE ENLARGEMENT).

In the pelletizing process, the iron ore must be ground to a very fine size (<75 μm(>200mesh)). The ground ore is mixed with the proper amount of water and binder, normally bentonite, hydrated lime, or organic material, and then is rolled into small balls 9–15 mm in diameter in a balling drum or disk. These green (wet) pellets are dried, then are heated to 1200–1375°C to bond the small particles, and finally are cooled. The heating can be done on a traveling grate, in a shaft furnace, or by a combination of a traveling grate and a rotary kiln (grate–kiln). The traveling grate and grate–kiln are the most commonly used pelletizing processes.

Sintering consists of igniting a mixture of iron-bearing limestone and coke fines on a traveling grate to produce a clinker-like aggregate (sinter) suitable for use in the blast furnace. The iron-bearing fines can include iron ore fines (sinter feed), iron ore concentrates, flue dust, or other steel mill wastes. The traveling grate is shaped like an endless loop of conveyor belt forming a shallow trough with small holes in the bottom. The bed of material on the grate is first ignited by passing under an ignition burner that is fired with natural gas and air; then, as the grate moves slowly toward the discharge end, air is pulled down through the bed. As the coke fines burn in the bed, the heat generated sinters the particles. At the discharge end of the machine, the sinter is crushed to remove extra large lumps, then cooled, and finally screened.

In the briquetting process, ore fines usually are mixed with a binder and are formed into compact masses between two rotating rolls. The rolls exert pressures of 1.5–4.2 t/cm² in forming the briquettes. In the nodulizing process, which is relatively uncommon commercially, ore fines are heated in a rotary kiln to a temperature, usually 1250–1370°C, at which the ore begins to melt and bind. The ore balls in the kiln to form nodules that are discharged and cooled.

Ironmaking Processes

Ironmaking refers to those processes which reduce iron oxides to iron. By the nature of the processes, the iron produced usually contains carbon and other impurities which are removed in downstream processing. There are three principal categories of ironmaking processes, in order of commercial importance: blast furnace, direct reduction, and direct smelting.

Blast Furnace. The blast furnace is the predominant method for making iron. Established for centuries as the premier ironmaking process, blast furnace ironmaking both enabled and profited from the Industrial Revolution. Although the fundamental principles of operation are unchanged, the blast furnace has evolved into a highly efficient and productive process.

In essence, the blast furnace is a large, countercurrent, chemical reactor in the form of a vertical shaft which is circular in cross section. Iron ore, coke, and fluxes constitute the burden which is charged continually into the top. Pressures in the shaft are controlled to 100–300 kPa (1–3 atms) gauge. Preheated air (hot blast) is blown in through water-cooled nozzles (tuyeres) around the circumference of the furnace near the bottom. The oxygen in the air reacts with the coke to form hot reducing gases (mostly carbon monoxide) which ascend through the burden and (1) provide heat for melting; (2) react with the iron ore to reduce it to iron; and (3) heat the ore, coke, and fluxes to reaction temperatures. Nitrogen in the hot blast is heated by the coke combustion, and aids in heat transfer to the burden. The gases leaving the top of the furnace (top gas) are cleaned, cooled, and used as fuel to preheat the air for the hot blast.

Molten iron (hot metal or pig iron) and slag (molten oxides) are produced and accumulate in the bottom of the furnace. The hot metal and slag are drained semicontinuously through a taphole (tapping, or casting) into a trough. The hot metal is separated from the slag by a weir–dam arrangement at the end of the trough, then flows through runners to a refractory-lined rail car. The hot metal is then transported to a nearby site for further processing. As shown in Table 4, about 99% of all pig iron produced in the United States is used for steelmaking. The remainder is cast into pigs for remelting or used directly for iron castings.

Table 4. U.S. Pig Iron Production and Consumption^a

Year	Production, t × 10 ⁶	Consumption for steelmaking	
		t × 10 ⁶	%
1983	48.706	45.783	94.0
1984	51.904	45.282	87.2
1985	50.446	49.547	98.2
1986	43.952	43.312	98.5
1987	48.410	47.413	97.9
1988	55.745	54.833	98.4
1989	55.873	55.299	99.0

1990	54.750	54.081	98.8
1991	48.637	48.154	99.0
1992	52.224	51.720	99.0

^a Ref. 9.

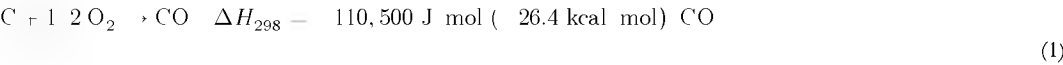
Raw Materials. Most of the iron enters the blast furnace as iron oxides, either hematite, Fe₂O₃, or magnetite, Fe₃O₄, in the form of pellets, sinter, or lump ore. Total iron content in the iron oxides normally ranges from 60 to 66%_o, but may be as low as 50%_o in low quality ores or in sinter using high quantities of recycled materials. Oxygen associated with the iron as iron oxide comprises about 23 to 28%_o. The remainder is gangue oxides, mostly silica and alumina. Silica should be less than 5%_o. Size control is important in order to promote permeability, so fines (< 6 mm) should be kept less than 2%_o. Pellets should be sized to at least 80%_o between 9 and 13 mm, with no more than 10%_o larger than 13 mm. Pellets are tested for handling characteristics by measuring tumble strength and compression strength. In the ASTM tumble test, 11.3 kg of pellets (screened to >6 mm) are placed in a drum 0.5 m wide by 1.0 m diameter. The drum is rotated at 25 rpm for 8 min, and the pellets are rescreened to measure degradation. Compression strength is a measure of the compressive force required to break the pellets between two parallel flat plates. For blast furnace usage, pellets should have a tumble strength of at least 95%_o >6 mm and a compression strength of greater than 200 kg. Other important properties include low swelling, little breakdown at low temperatures during reduction, and good reducibility (10). In some locations, steelmaking slag and yard scrap are crushed and magnetically upgraded to recycle iron units through the blast furnace. Metallic iron, usually in the form of direct reduced iron (DRI) or hot briquetted iron (HBI), is also sometimes added to increase productivity and lower fuel rate. The overall feed mix of iron units consumed in the U.S. blast furnaces in 1992 consisted of 79%_o pellets, 15%_o sinter, 2.9%_o lump ore, 3%_o ferrous scrap, and 0.1%_o HBI (9).

Carbon is provided by the coke, which typically consists of 85–90%_o fixed carbon, less than 2%_o volatile matter, 5–13%_o ash, 0.6–1.3%_o sulfur, and 2–10%_o moisture. The carbon is required as the reductant, and in addition provides heat through combustion with air. Coke is the preferred form of carbon because it provides structural support within the furnace by creating stable areas of permeability for the ascending gases, especially in the softening melting zone. High mechanical strength of coke is important for smooth operations. Coke is most often sized to between 15 and 75 mm, with many operators specifying a more narrow range. In some cases, coke is screened to two or more different size fractions for separate charging (see COAL CONVERSION PROCESSES, CARBONIZATION).

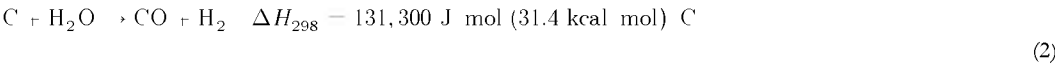
Fluxes are usually added in the form of either limestone or dolomite. The fluxes provide the basic constituents (CaO and MgO) needed to balance the acid constituents (SiO₂ and Al₂O₃) from the coke and ore. These are the four primary oxides which form the slag, although minor amounts of other oxides such as MnO, Na₂O, K₂O, P₂O₅, and TiO₂, and sulfur are also present. Proper adjustment of the slag chemistry is necessary to obtain the desired balance of chemical properties, eg, desulfurization and alkali removal, and physical properties, such as low melting point and low viscosity. There has been a growing practice of adding CaO and MgO at the pelletizing process to make fluxed pellets. The advantages of fluxed pellets go beyond simply avoiding the thermal requirement of calcining limestone and dolomite (heating to drive off volatile matter) (11,12). Pellet properties are improved, which results in improved blast furnace performance.

Air for the hot blast may also be considered a raw material. The air is preheated in stoves to between 900 and 1300°C. Over 1.5 t of air is required to produce 1 t of hot metal (pig iron). Solid, liquid, or gaseous fuels, eg, coal, fuel oil, or natural gas, may be added to the hot blast at the tuyeres to replace some of the coke. Oxygen may also be added to the hot blast to increase flame temperature.

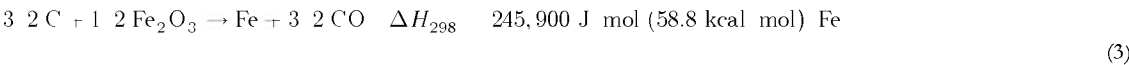
Thermochemistry. From an overall heat and mass balance point of view, the main chemical reactions of the blast furnace include oxidation of carbon in the zone in front of the tuyeres (raceway) to give CO plus heat,



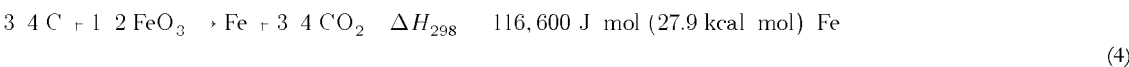
reduction of the moisture in the hot blast by carbon to form CO plus hydrogen,



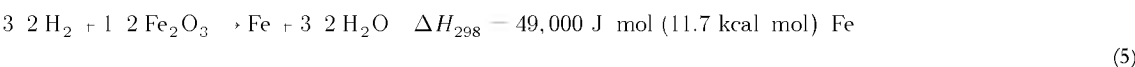
reduction of iron oxide by carbon to form iron and CO,



reduction of iron oxide by carbon to form iron and CO₂,

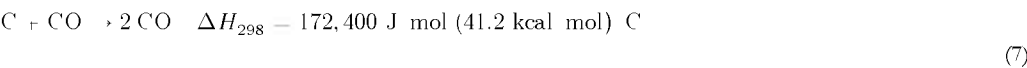
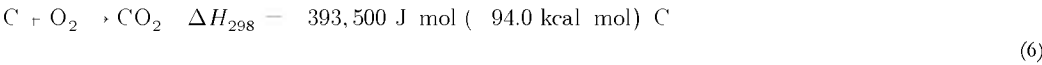


and reduction of iron oxide by hydrogen to form iron and H₂O.



All the reduction reactions are endothermic, regardless of the reductant used. The heat for these reactions, along with the requirements for the sensible heats of the hot metal and slag, and heat losses through the furnace shell, is provided by the heat generated from equation 1 plus the sensible heat of the hot blast.

The furnace (Fig. 2) may be divided into four zones (from bottom to top). (1) Hearth and raceway: as the coke descends through the furnace, it is heated by the ascending gases to about 1370°C. When it reaches the raceway in front of the tuyeres, it reacts immediately with the oxygen in the hot blast air. Equation 1, however, is actually the combination of coke combustion (eq. 6) and coke gasification (eq. 7, also referred to as solution loss).



Coke gasification occurs just outside the raceway area where gaseous oxygen is no longer available to completely combust the CO to CO₂. This reaction goes essentially to completion at temperatures between 1500 to 2100°C. The net heat effect is exothermic, as shown in equation 1. The endothermic equation (eq. 2) allows control of the temperature in front of the tuyeres by controlling the moisture in the hot blast.

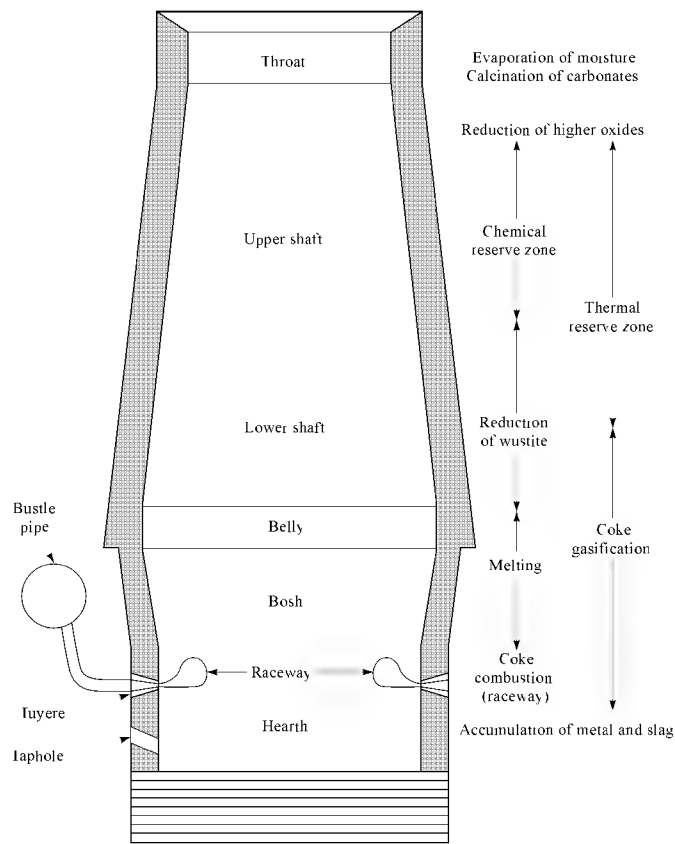
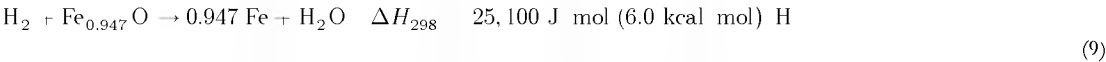
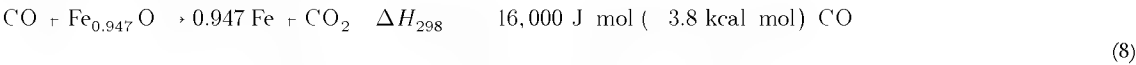


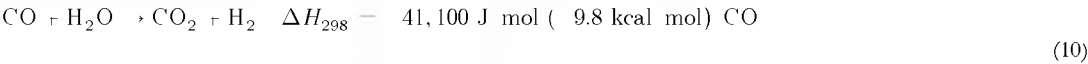
Fig. 2. Schematic of a blast furnace.

(2) Melting (fusion) zone and final reduction of wustite: the H_2 and CO rise through the burden, contact wustite [17125-56-3] formed from previous reduction reactions in the upper part of the furnace, and reduce it to iron.

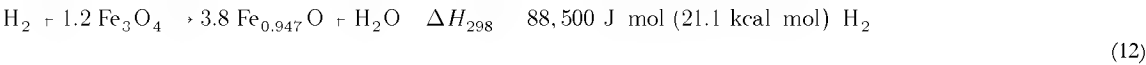
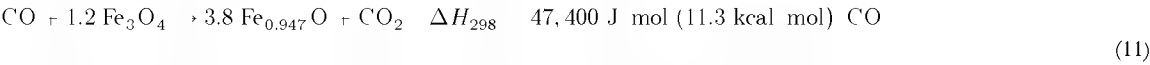


The iron absorbs carbon through contact with the coke, which melts owing to its decreased melting point. Equation combines with equations 8 and 9 in a cycle which effectively regenerates CO . Owing to the highly endothermic nature of equation 7, the gases cool as they rise in the furnace. Equation , reduction by CO , is referred to as indirect reduction. The combination of equations 7 and 8, solution loss and indirect reduction, is referred to as direct reduction, because it amounts to reduction of the wustite directly by carbon to form iron and CO . This direct reduction is not the same terminology used in direct reduction processes, which in fact often rely on indirect reduction reactions (see IRON BY DIRECT REDUCTION).

(3) Thermal reserve zone: once the gases (CO , H_2 , and N_2) have cooled to about $925^\circ C$, the thermodynamics for equation 7 are no longer favorable. Because the predominant reaction is now equation 8 which is slightly exothermic, and because the mildly endothermic equation 9 occurs to a much lesser extent, the gases do not cool appreciably, resulting in a thermal reserve zone. The net relative amounts of CO_2 and H_2O produced by reduction are determined by equilibrium for the water gas shift reaction,



(4) Reduction of magnetite and hematite (upper shaft): as more and more CO and H_2 are converted to CO_2 and H_2O , the ascending gases eventually become too weak in concentration to reduce wustite to iron. Figure 3a and b shows the regions of stability for the $Fe-C-O$ and the $Fe-H-O$ systems, respectively. The boundary lines represent the equilibrium conditions for the various reduction reactions as shown. The gases, too weak to reduce wustite, are strong enough to reduce magnetite to wustite:



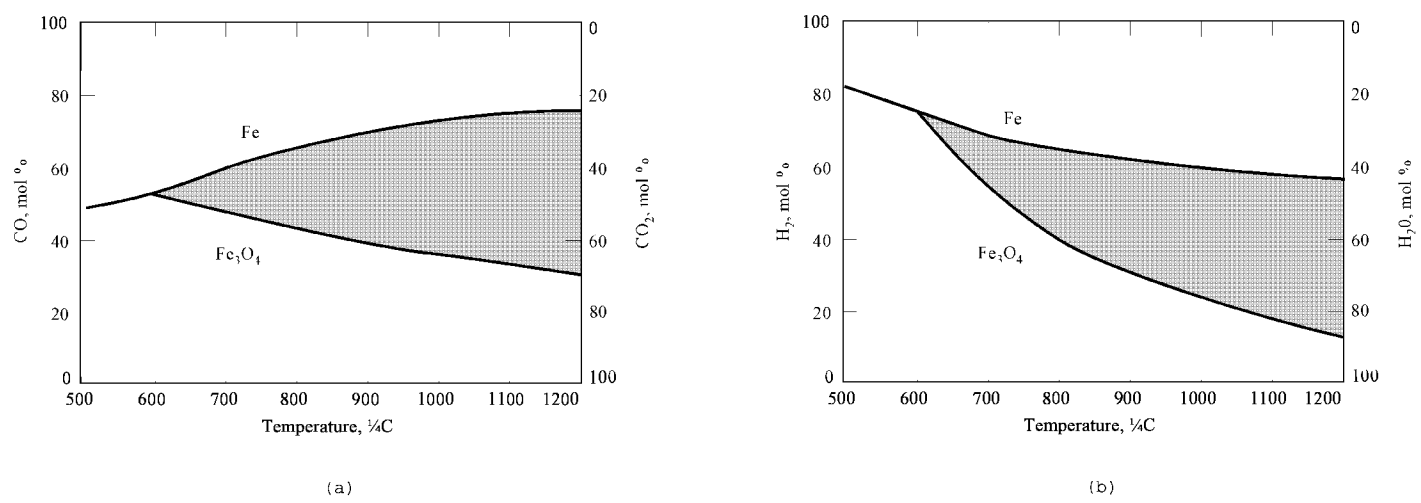
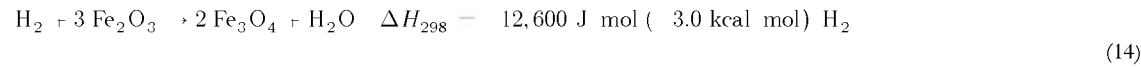
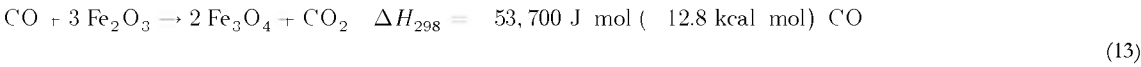


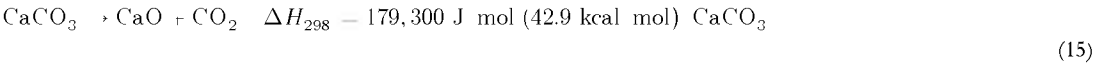
Fig. 3. Phase diagrams in which (▨) represents FeO. (a) Fe–O–C system; (b) Fe–O–H system.

For steady-state operation, the amount of wustite produced by these reactions must exactly match the amount of wustite reduced to iron in the lower part of the furnace. Owing to equilibrium and stoichiometric considerations, there is more than enough CO generated from the wustite–iron reactions. This results in driving the region of reduction of magnetite toward the top of the furnace, and creates a chemical reserve zone of little reaction between the descending wustite and the weakly reducing (with respect to wustite) ascending gases. Because there is little heat required for this zone, the chemical reserve zone coincides with the upper part of the thermal reserve zone. The ascending gases rapidly reduce the hematite which is charged into the top of the furnace.

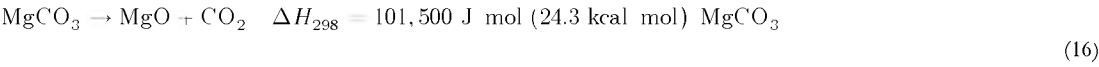


Only the slightest amounts of CO or H₂ are required to reduce hematite to magnetite, which is why in Figure 3 the regions of magnetite stability are shown extending all the way down to the bottoms of the graphs. Owing to the excess CO generated by the wustite reduction reactions, hematite reduction is also driven to the top of the furnace. The reduction of hematite to magnetite and magnetite to wustite is so fast that hematite, magnetite, and wustite may all be found in the same pellet, owing to the topochemical (occurring at boundaries which progress from surface to center) nature of the reactions. In this zone the gas temperature falls off rapidly because of cooling by the incoming materials, evaporation of moisture, and the net endothermic nature of the reduction reactions.

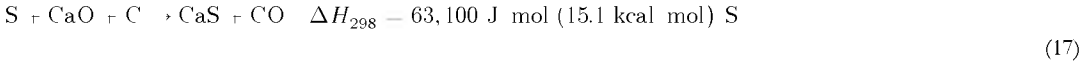
In addition to the principal equations discussed, several others occur which may be of importance, including calcination of calcium carbonate (limestone), which takes place in the upper shaft at 800–870°C,



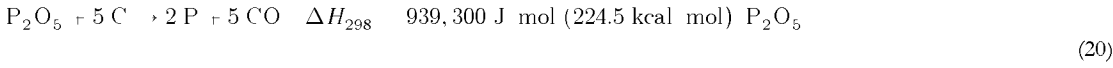
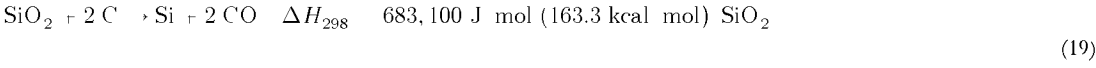
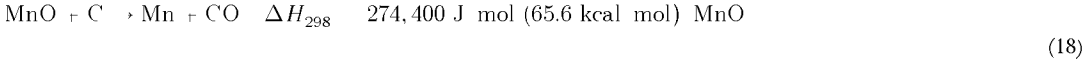
calcination of magnesium carbonate in dolomite which takes place in the top of the shaft at about 50–100°C,



fluxing of the sulfur into the slag,



and reduction of other metallic oxides,



Equations 17–20 result from contact between hot metal and slag, and the sulfur and carbon come dissolved in the hot metal. Likewise, the manganese, silicon, and phosphorus which are produced are dissolved into the hot metal. The heats of solution for these elements in some cases depend on concentration, and are not included in the heats of reaction listed above. The ratio of the concentration of the oxide (or element for sulfur) in the slag to the concentration of the element in the hot metal is the partition ratio, and is primarily a function of slag chemistry and temperature.

Mass and energy balances are used to evaluate blast furnace performance. Many companies now use sophisticated computerized data acquisition and analysis systems to automatically gather the required data for daily calculation of the mass and heat balances. Typical mass and heat balances are shown in Figure 4 and Table 5, respectively.

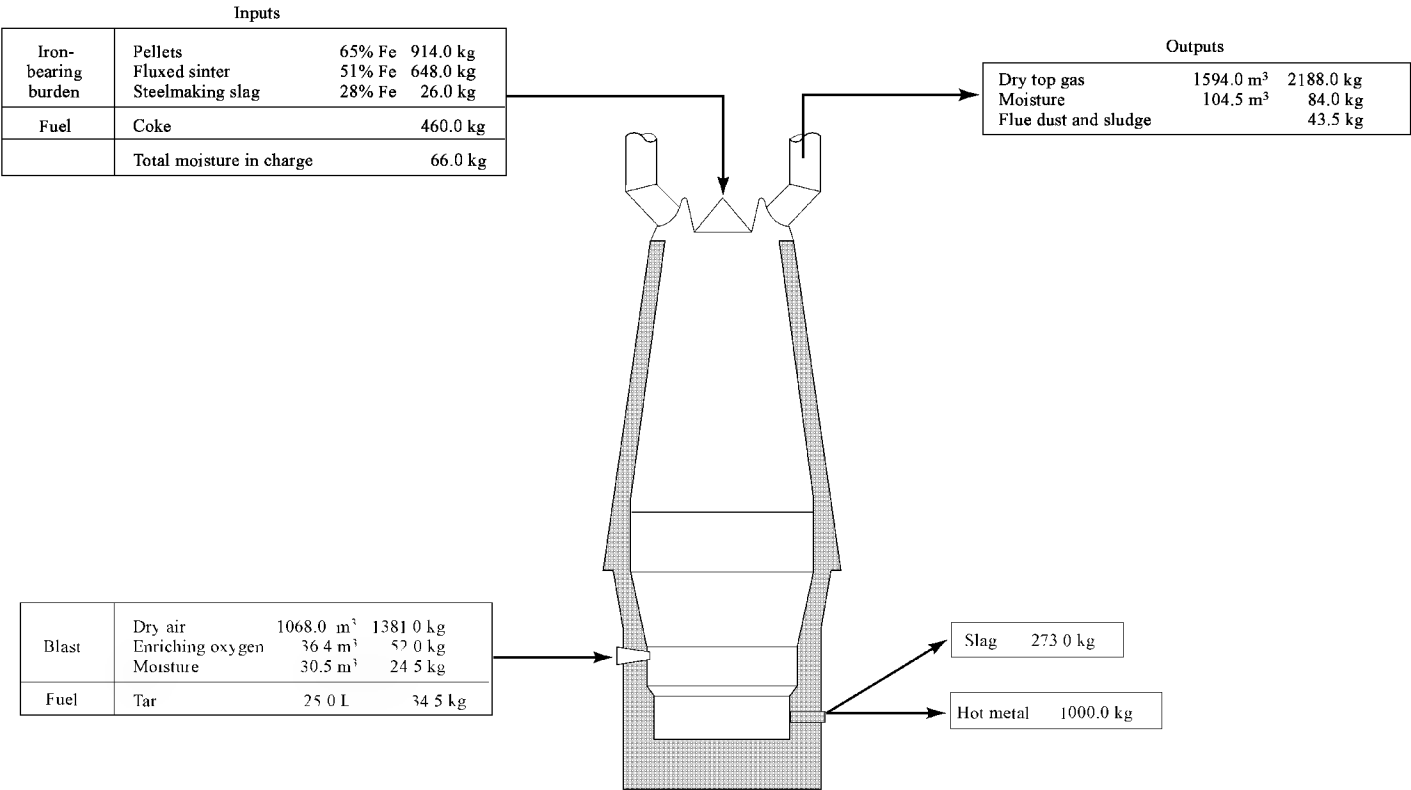


Fig. 4. Blast furnace material balance where all quantities represent the amount per metric ton of hot metal. The dry top gas contains 23.4% CO, 21.2% CO₂, and 2.5% H₂; the slag contains 38% SiO₂, 9% Al₂O₃, 42% CaO, 10% MgO, and 1.3% S; and the hot metal contains 4.5% C, 0.48% Si, 0.59% Mn, 0.029% S, and 0.060% P (1).

Table 5. Blast Furnace Energy Balance^a

Balance	Hot metal, GJ t ^b	%
Energy input		
sensible heat of hot blast	1.64	41
combustion of coke	2.10	52
of injected fuel	0.27	7
Total	4.01	100
Energy output		
reduction of iron oxides	1.17	29
of other metalloids	0.15	4
sensible heat of slag	0.48	12
of hot metal	1.36	34
of top gas	0.20	5
decomposition of H ₂ O at raceway	0.26	6
vaporization of H ₂ O from burden	0.13	3
heat losses	0.26	6
Total	4.01	100

^a Ref. 1.
^b To convert GJ to Btu, multiply by 0.9488 × 10¹².

Plant Layout. Figure 5 shows the material flow diagram for a blast furnace plant. The ore and fluxes are stockpiled in a large open yard, from which these are reclaimed by crane (ore bridge) and transferred by conveyor to the stockhouse. Coke is delivered by rail or conveyor from the coke plant. The stockhouse consists of a row of bins from which the raw materials are weighed out in the desired order and amount, and conveyed by either skip car or conveyor to the top of the furnace. Special additives to the charge, such as upgraded yard scrap, manganese ore, or calcium chloride, for flushing out alkalis, are added at this location.

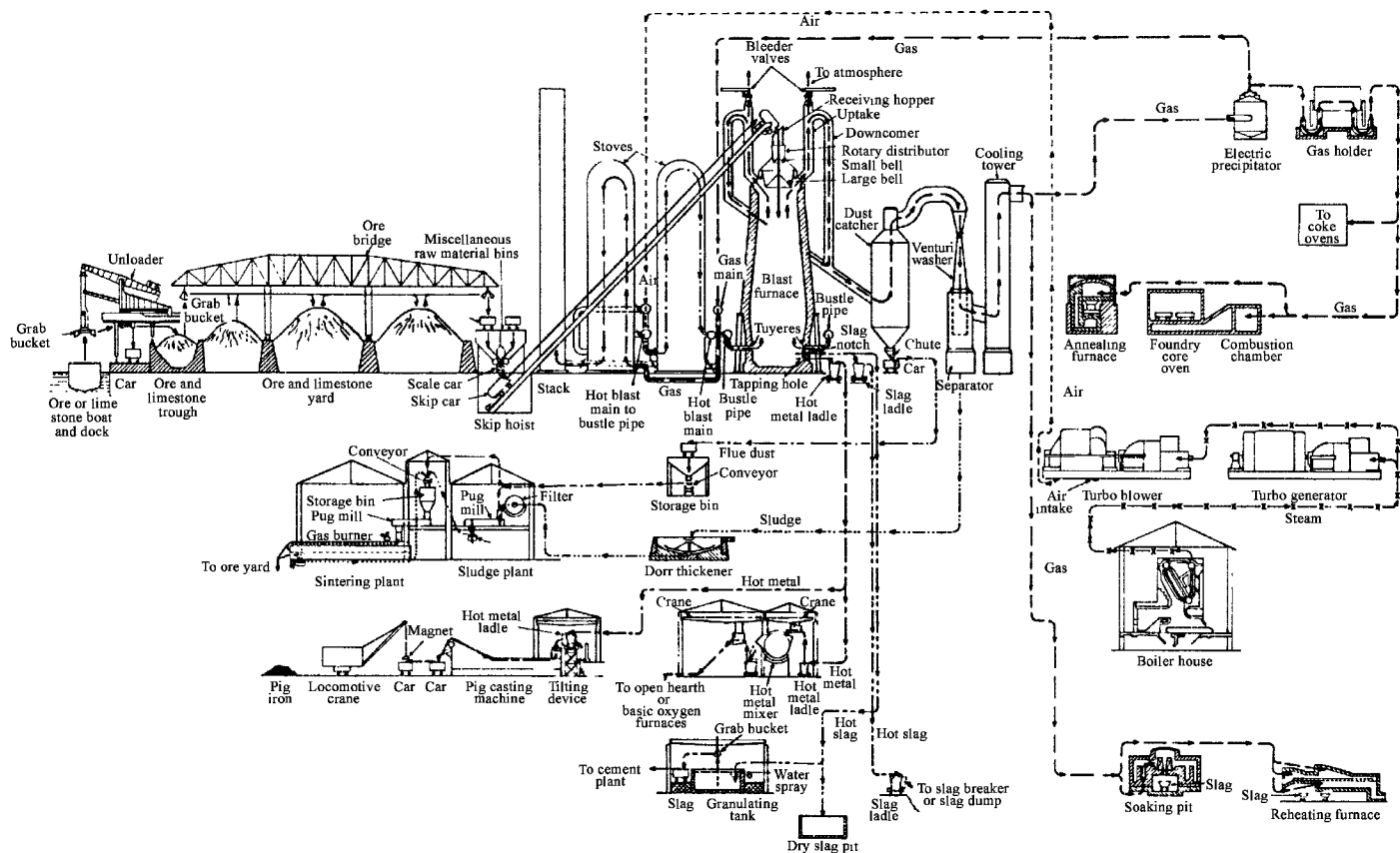


Fig. 5. Flow diagram depicting the principal units and auxiliaries in a modern blast furnace plant, and showing the steps in the manufacture of pig iron from receipt of raw materials to disposal of pig iron and slag, as well as the methods for utilizing the furnace gases. (· · ·), Miscellaneous raw material; (- - -), cold-blast air; (· · ·), hot-blast air; (———), blast furnace gas; (— · —), steam; (— — —), hot metal; (- - - -), hot slag; (·····) flue dust; (- · - · -), sludge; and (···), sinter.

At the top of the furnace, the raw materials are charged into the furnace through a pressurized gas seal system, typically a double bell (see Fig. 6a). For very high pressure furnaces, three bells may be used. More recently, the Paul Wurth bell-less top has become popular, as its rotating chute design permits greater control over distribution of the burden (Fig. 6b).

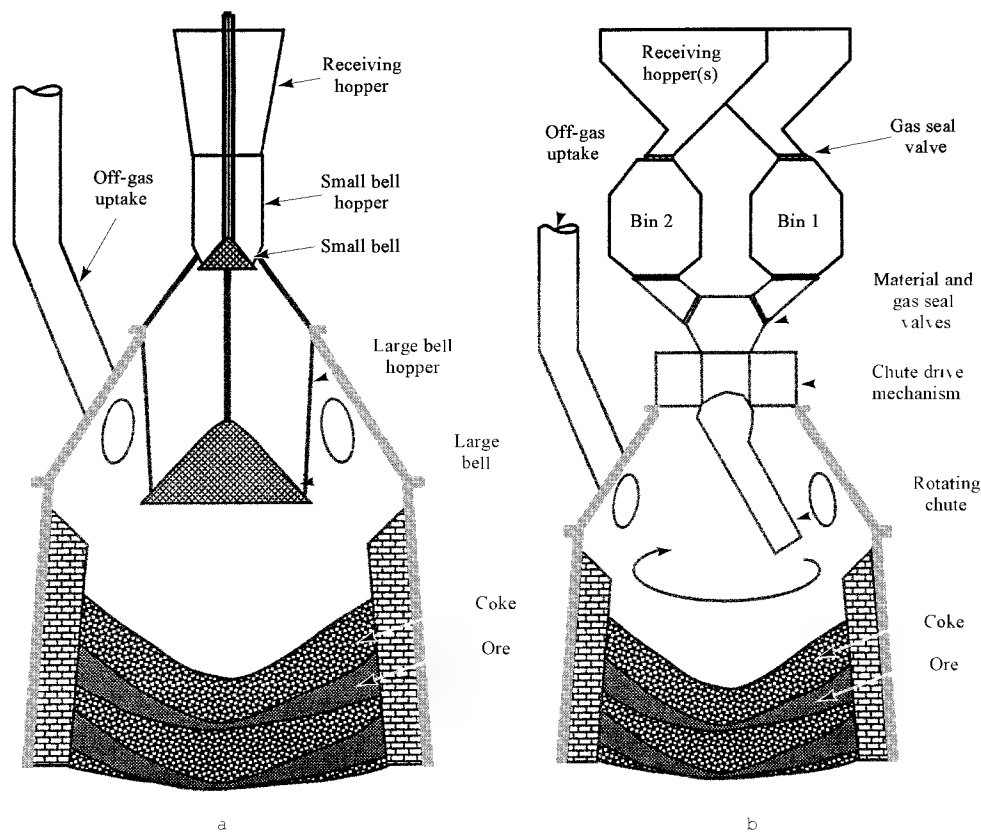


Fig. 6. Pressurized gas seal system at top of blast furnace: (a) two-bell top; and (b) bell-less top.

The taphole is built into the refractory lining of the blast furnace. The taphole drill is used to drill a hole through the taphole material. At the end of the cast, a mud gun is used to plug the hole with a quick-hardening clay. An alternative technique (hot bar) relies on hardening the refractory around a metal bar, which is pulled out for tapping. Hot metal is tapped (cast) every three to five hours into refractory-lined railcars for transportation to nearby steelmaking facilities or to a pigging machine. The largest blast furnaces have up to four tapholes which are used alternately as the trough and runners are repaired. Slag is either transferred as a liquid in inverted bell-shaped rail cars, or poured directly into a slag pit adjacent to the blast furnace. After solidification, the slag is crushed and sized and sold for road ballast. In addition, the slag may be granulated using a water spray to make a by-product

suitable for sale to the cement (qv) industry, or used for the production of rock wool insulation.

Off-gases (top gas) leave the top of the furnace through uptake pipes, reverse direction in the downcomer, and enter the dust catcher, in which condensed water and dust are separated from the gases. The wet dust is emptied into a rail car for transport to a sinter plant for recycle or to a landfill. After leaving the dust catcher, the off-gas is washed in a venturi scrubber. The cleaned gas is then used within the plant for steam (qv) generation, to fire annealing or other furnaces, for underfiring of coke ovens, and for firing the blast furnace stoves. There are usually three or four stoves filled with refractory checkers. The checkerwork contains a multiplicity of passages through which gases may pass. The stoves alternate between absorbing heat generated by combustion of the blast furnace off-gas and releasing heat to the cold blast air as it passes down through the heated checkers.

After it leaves the stoves, the hot blast enters a large refractory-lined bustle pipe to distribute the gas evenly around the furnace. Multiple connecting pipes (tuyere stock) direct the hot blast to the blowpipes. At the ends of the blowpipes are the tuyeres, water-cooled copper nozzles set into the refractory lining of the blast furnace.

Operation. Because of the long residence time of the materials (8–10 h), the blast furnace process can exhibit considerable inertia, and control is usually applied where the goal is maintaining smooth, stable input conditions. One of the most important aspects of blast furnace control is supply of consistent quality raw materials, which is why there is a strong emphasis on quality control at coke plants, pelletizing plants, and sinter plants (see [QUALITY ASSURANCE](#), [QUALITY CONTROL](#)).

Careful attention is paid to the properties of the raw materials. Short-term adjustments to the slag basicity may be made by increasing or decreasing the amount of fluxes. In general, increasing the basicity of the slag promotes sulfur removal but decreases alkali removal. A balance is sought between the two, in addition to ensuring that the slag has a low melting point and low viscosity for good fluidity. Longer term adjustments may be made by adjusting the chemistry of the pellets and the sinter.

Modern furnaces rely on computerized controls for weighing and charging of the raw materials. The ore and coke are charged in alternating batches, so as to create distinct layers within the furnace which promote permeability for the rising gases. For a two-bell top, the manner in which the layers are formed is a function of the trajectory of the burden materials as these fall off the bell and rebound from the throat. Control of the layering is exercised by carefully selecting the amounts and order of materials. Sophisticated physical and computer models are used to guide the operator in deciding exactly how much and in what order the raw materials should be charged.

Moveable throat armor is used for additional control. Hanging plates or horizontally adjustable guides are used to adjust the throat diameter for each charge, thus modifying how the layers are formed in the furnace. Bell-less tops (Fig. 6b) provide even greater flexibility, as both the angle and the rotation speed of the rotating chute may be adjusted. Many furnaces use probes, radar, or laser devices to provide feedback on burden distribution.

Computer controls are likewise used for stove operation, to control delivery of the hot blast. High hot blast temperatures are generally desirable, as these reduce the coke rate. Control of the flame temperature in the raceway is effected by controlled additions to the hot blast, primarily of moisture. Injectants into the tuyeres such as coal, oil, and natural gas are often used to replace some of the coke. The effect of these injectants on flame temperature must be accounted for, and compensation is performed by lowering moisture or adding oxygen.

Tapping or casting is controlled to avoid rapid drops in the burden. Normally, the hot metal accumulates in the hearth for some period of time, then is drained during tapping. Proper selection of taphole, trough, and runner materials, improved trough and dam designs, and multiple tapholes have all allowed the proportion of time spent tapping to increase to the point where it is nearly continuous. Improved taphole clays have reduced erosion of the taphole, stabilizing the casting rate and promoting smooth descent of the burden. Careful raw materials selection, burden distribution (charging control), hot blast temperature and moisture control, tuyere injectants, and casting practices are the primary means by which the blast furnace operator keeps the ironmaking process under control.

Direct Reduction. Direct reduction processes are distinguished from other ironmaking processes in that iron oxide is converted to metallic iron without melting. Because this product, called direct reduced iron (DRI), is solid, it is most suitable for melting in an electric arc furnace (EAF) as a substitute for scrap (see [FURNACES, ELECTRIC](#)). The briquetted form of DRI, hot briquetted iron (HBI) is used when the product is to be transported. Briquetting increases density and chemical stability. The predominant direct reduction processes (MIDREX and HyL III) are based on natural gas as a fuel and reductant source. They are economically attractive in regions where natural gas is cheap and abundant, especially if iron ore is available nearby (see [IRON BY DIRECT REDUCTION](#)).

Direct Smelting. Direct smelting processes use coal directly instead of coke. Several processes are under development which effectively divide the functions of the blast furnace into two separate but connected unit operations. First, the iron ore is prereduced in a shaft furnace or a fluidized bed, depending on the process and the type of ore used. Second, the prereduced ore is charged into a molten bath into which coal and oxygen or air are also introduced. The gases leaving the smelter are used to perform the reduction in the prereduction vessel.

The Corex process is the only one of the newer ironmaking processes operating on a commercial scale. There is a 350,000 t/yr plant in South Africa and a 700,000 t/yr plant being constructed in Korea. In the Corex process, over 90% of the reduction is performed in the prereduction shaft; the remainder is accomplished in what is primarily a melter–gasifier. Oxygen and coal are injected into the melter–gasifier to provide heat for melting and to generate the reducing gases for the prereduction furnace. Because of the high proportion of reduction performed in the prereducer, the off-gases leaving the melter–gasifier are necessarily highly reducing. As a consequence, the fuel rate for the melter–gasifier is higher than for processes in which the gases are more completely oxidized. This is compensated for to some extent by the low heat load required for reduction. However, the net effect is that relatively high volumes of off-gas having significant chemical heat value are generated which require large off-gas handling systems. This may be an advantage in locations where the energy value of the off-gas can be utilized fully to replace other energy sources such as natural gas or electricity.

The processes being developed by the American Iron and Steel Institute and the Department of Energy (AISI-DOE) in the United States and the Japan Iron and Steel Federation (JISF) in Japan share similar features in the smelter, but differ in prereduction approaches. In the AISI-DOE process, pellets are prereduced to wustite, about 30% prereduction, in a shaft furnace. In the JISF process, called direct iron oxygen smelting (DIOS), iron ore fines are prereduced to between 30 and 60% in one or more fluidized beds in series. For both, the prereduced ore and coal are charged into a vertical vessel containing a molten bath, and oxygen is injected to generate CO and heat. Additional oxygen is provided to post-combust the CO to CO₂, thereby improving the energy efficiency of the process.

The ore melts to a liquid oxide almost immediately upon introduction into the smelter. The primary reduction reactions, then, are between the liquid oxide and carbon-saturated liquid iron or solid particles of char contained in the slag. The generation of CO within the bath results in a foamy slag. Control of slag foaming, high post-combustion levels, and high post-combustion heat-transfer efficiency are critical operating factors for these processes. Both processes are under development in the pilot stage (250–500 t/d). Commercial development is envisioned for the year 2000.

The HiSmelt process being developed jointly by CRA of Australia and Midrex Direct Reduction Corp. uses a horizontal vessel, relying on turbulence in the bath to spray particles of slag and iron into the atmosphere above the bath, where heat is transferred from the post-combustion flame to the particles. Here, air is used instead of oxygen, thus removing the requirement of an oxygen plant. This technology emphasizes bottom injection of coal and dust into the iron bath.

Production and Economics

As shown in Figure 7, pig iron production in the United States rose dramatically from about 10 × 10⁶ t/yr in the mid-1890s to about 90 × 10⁶ t/yr in 1970. The negative impact of the Depression of the 1930s and the iron demands of World War II in the 1940s are both visible. The strong overall trend up to the 1970s led to many steel companies investing in partnership positions with iron ore mining companies and making long-term take-or-pay contracts. However, the oil embargo in 1973 resulted in a serious decline in steel demand and concomitantly in pig iron production in the United States. In addition, the rapid expansion in pig iron production in the rest of the world (Fig. 8) limited the opportunities for export.

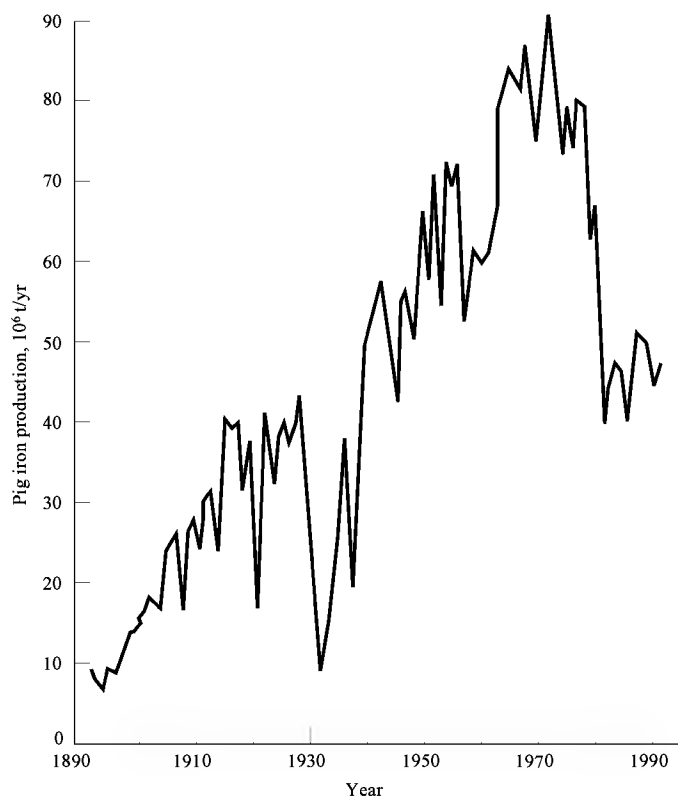


Fig. 7. U.S. pig iron production (13,14).

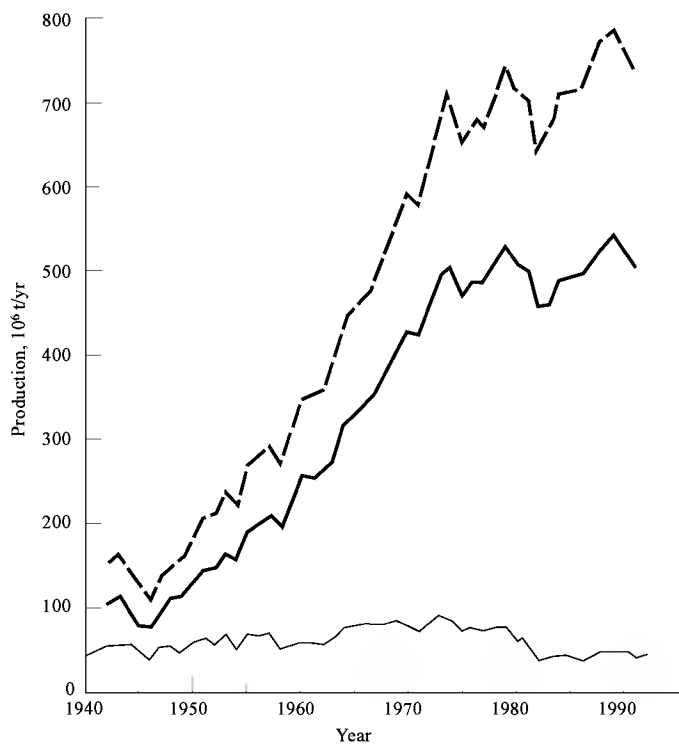


Fig. 8. Production of (—) U.S. and (- - -) world pig iron and (. . .) world steel (13,14).

As can be seen in Figure 8, the proportion of world pig iron produced in the United States has decreased dramatically since 1950. Also notable is the widening gap between pig iron and steel production, indicating the increasing use of recycled iron or scrap (see RECYCLING, FERROUS METALS) and alternative iron sources such as DRI and HBI. The increased demand for scrap is reflected in scrap iron prices (Fig. 9), which in turn have spurred growth in direct reduction processes.

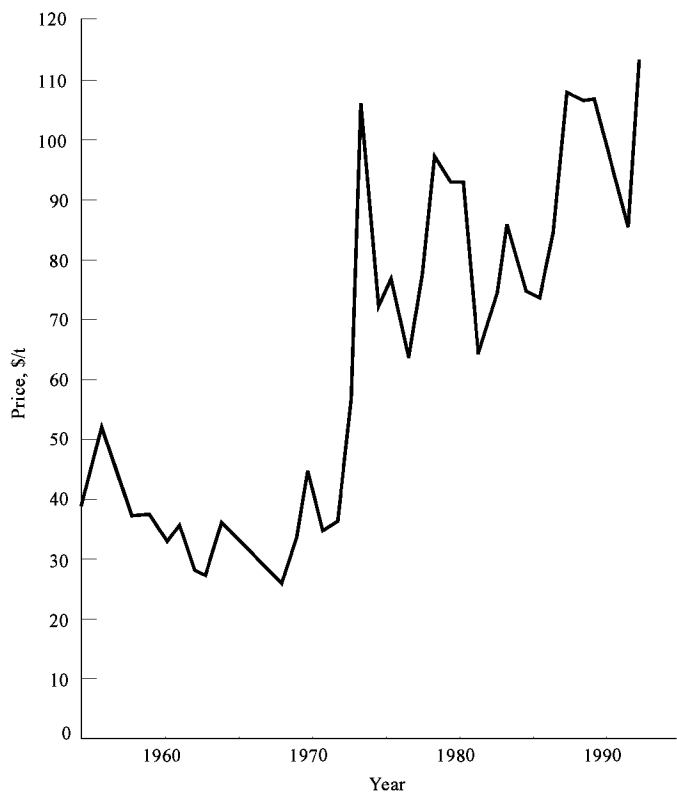


Fig. 9. Average prices for scrap. No. 1 heavy metal, composite averages of Chicago, Pittsburgh, and Philadelphia (13,14).

Historically, the economies-of-scale involved in blast furnace ironmaking led to the replacement of numerous small furnaces by larger units. The largest blast furnaces of the 1990s are over 14 m in diameter at the hearth, and produce over 10,000 t d. However, economic conditions have made the massive capital requirements of such large furnaces too great to manage. To minimize the high costs associated with relining, improved water cooling and refractory systems have been used to extend blast furnace campaigns to 10 years or more. Additionally, environmental restrictions on coke plants have increased costs and made new investments in cokemaking nearly prohibitive. Coke batteries have a limited life of about 30 years, and many plants in the United States, Japan, and Europe will be close to needing replacement in the first decade of the twenty-first century.

These factors have prompted two principal thrusts in ironmaking development. First, progress continues to be made in increasing blast furnace productivity and in decreasing coke rates. Coal (qv) injection to replace coke units has assumed a prominent role. Coal replaces coke on a nearly 1:1 mass basis, and coal injection rates of up to 250 kg t of hot metal (thm) have been achieved. Injection of oxygen and other reductants besides coal are expected to be used more extensively. Increased additions of scrap, DRI, and HBI are expected to play a significant role in efforts to boost productivity and decrease coke rates.

Second, development efforts in direct reduction and direct smelting processes have also increased. Whereas these developments require significant commitment of finances and labor to achieve commercialization, the rewards promised by avoiding cokemaking and by utilizing smaller scale units which have relatively low capital costs make the commitment worthwhile.

Globally, iron production is expected to increase in developing countries as local steel industries grow to supply the increasing demand for steel products. Iron production in already developed countries is expected to stabilize or possibly decline as the opportunities for export diminish. Efforts in the developed countries are expected to be in energy efficiency, productivity, quality, and cost reduction.

Environmental Issues

Tremendous progress was made in the 1980s and 1990s in response to environmental issues, especially in the area of emissions from ironmaking facilities. Dust is controlled by protecting open piles, watering roadways, covering conveyor belts and transfer points, controlling fumes through improved casthouse practices, and air cleaning systems ducted to baghouse or other filtration systems (see AIR POLLUTION CONTROL METHODs). Sulfur dioxide emissions are controlled using off-gas and stack cleaning systems (see SULFUR REMOVAL AND RECOVERY).

Mining practices have been altered to include reclamation of areas where open pit mining has occurred. Safety practices throughout the ironmaking processes are continually being upgraded through training, improved operating practices, and installation of sophisticated detectors and automatic shutdown systems.

Perhaps the biggest environmental challenge for ironmaking processes into the twenty-first century involves responding to the concerns about global warming. Ironmaking processes require the use of carbon-based reductants, and ultimately result in the emission of carbon dioxide.

Cast-Iron Production

Most ferrous scrap is recycled in steelmaking processes by melting the scrap in either a basic oxygen or an electric arc furnace. However, a significant market exists for cast-iron products, which are also made by melting ferrous scrap. In 1991, world production of cast irons was estimated at nearly 3.9 × 10⁷ t at over 14,000 iron foundries (15).

Cast irons are normally produced by melting iron or steel scrap along with pig iron. The carbon and silicon levels are adjusted to obtain the desired properties. Melting is done in cupolas, electric furnaces, or air furnaces. The cupola resembles a small blast furnace, but differs in that pig iron and scrap replace the ore. Coke combustion using air provides the heat for melting of the charge. Electric arc furnaces are used to a limited extent, but induction furnaces are more popular. The air furnace is a type of reverberatory furnace which has a fireplace at one end, a stack at the other end, and a hearth in between. The cupola is the most common source of iron for casting. Cast irons may be classified as either gray, white, malleable, or ductile iron. Silicon and carbon are the two most important elements used to adjust the compositions of cast irons, but other elements such as Mn, S, P, Cr, Ni, Mo, and Cu may also be important. The primary effect of these elements is on the form the carbon takes as it precipitates during solidification. As shown in Figure 1, cast irons when solidified are mixtures of alpha iron plus graphite or iron carbide, Fe₃C, or cementite.

Gray iron contains most of its carbon in graphite form as flakes. White iron has lower levels of carbon and silicon, resulting in nearly all the carbon in the form of iron carbide. Malleable irons are produced by heat treating white iron such that the carbon diffuses from the iron carbide to form graphite in a roughly spheroidal shape. It is the shape of the graphite that permits the increased malleability of the product. Ductile (or nodular) iron is produced by adding cerium or magnesium to iron having slightly higher carbon and silicon but lower sulfur than gray iron. The addition of these special agents also promotes the formation of spheroidal graphite. The success of the foundry industry results in part from its versatility. Castings can often be more intricate

than would be obtainable by machining. A wide range of properties, from brittle to ductile, may be obtained. Castings may weigh as little as a few grams or several tons.

Health and Safety

Iron presents minimal health risks. Skin contact should not result in any adverse health effect. Excessive inhalation of dust may be irritating to the respiratory tract. Dust may also cause mechanical irritation on eye contact. Extremely large oral doses would be required to cause gastrointestinal disturbance. The LD₅₀ toxicity rating (RTECS, 1992) for oral ingestion is 30 g kg. Iron is not categorized as hazardous or limited under any of the following regulations: SARA Sec. 302 EHS (RQ or TPQ), SARA Sec. 313 Chemicals, CERCLA Sec. 103 RQ, or RCRA Sec. 261.33 (16).
Iron dust does present a moderate fire and explosion hazard when exposed to heat and flame. Although normally not very reactive, under certain circumstances iron can react with water to liberate flammable hydrogen gas.

BIBLIOGRAPHY

"Iron" in *ECT* 1st ed., Vol. 8, pp. 18–45, by W. Carroll, Republic Steel Corp.; in *ECT* 2nd ed., Vol. 12, pp. 1–21, by W. A. Knepper, United States Steel Corp.; in *ECT* 3rd ed., Vol. 13, pp. 735–753, by W. A. Knepper, United States Steel Corp.

1. W. T. Lankford, Jr. and co-workers, *The Making, Shaping, and Treating of Steel*, 10th ed., Association of Iron and Steel Engineers, Pittsburgh, Pa., 1985.
2. G. S. Brady and H. R. Clauser, *Materials Handbook*, 11th ed., McGraw-Hill Book Co., Inc., New York, 1977.
3. R. M. Brick and A. Phillips, *Structure and Properties of Alloys*, McGraw-Hill Book Co., Inc., New York, 1949.
4. M. Sherwood, "Chemistry Today," in *The World Book Encyclopedia of Science*, Vol. 3, World Book, Inc., Chicago, 1989.
5. S. L. Goodale, *Chronology of Iron and Steel*, Penton Publishing Co., Cleveland, Ohio, 1931.
6. *Mineral Commodity Summaries 1993*, U.S. Bureau of Mines, Washington, D.C., 1993.
7. *Iron Ore Statistics 1993 Supplement*, United Nations Conference on Trade and Development, Geneva, Switzerland, 1993.
8. F. D. DeVaney, "Iron Ore," in *SME Mineral Processing Handbook*, Society of Mining Engineers of AIME, New York, 1985.
9. *Annual Statistical Report*, American Iron and Steel Institute, Washington, D.C., 1992.
10. J. J. Poveromo, "AISI Survey of Blast Furnace Pellet Properties," in *Proceedings, 52nd Ironmaking Conference*, AIME-ISS, Dallas, Tex., 1993.
11. C. K. Russell and co-workers, in Ref. 10.
12. P. C. Chaubal and co-workers, in Ref. 10.
13. *Metal Statistics 1971*, 64th ed., The American Metal Market, Somerset, N.J., 1971.
14. *Metal Statistics 1988*, 64th ed., The American Metal Market, Fairchild Publications, New York, 1988.
15. "26th Census of World Casting Production – 1991", *Mod. Cast. Mag.* (Dec. 1992).
16. *Material Safety Data Sheet, Reduced Iron*, Mallinkrodt Specialty Chemicals Co., Paris, Ky., Jan. 1993.

General References

G. S. Brady and H. R. Clauser, *Materials Handbook*, 11th ed., McGraw-Hill Book Co., Inc., New York, 1977.
J. G. Peacey and W. G. Davenport, *The Iron Blast Furnace*, Pergamon Press, London, 1979.
Iron Ore 1992, American Iron Ore Association, Cleveland, Ohio, 1992.
D. M. Kundrat, handouts from *Modeling in Process Metallurgy*, University of Cincinnati, Ohio, 1985.

J. A. Lepinski
PT Perkasa Indobaja

Jeffrey C. Myers
Midrex Direct Reduction Corporation

IRON BY DIRECT REDUCTION

Direct reduction (DR) is the process of converting iron ore (iron oxide) into metallic iron without melting. The metallic iron product, known as direct reduced iron (DRI), is used as a high quality feed material in steelmaking.

The most common method of converting iron ore to metallic iron utilizes a blast furnace wherein the material is melted to form hot metal (pig iron). Approximately 96% of the world's iron is produced this way (see IRON). However, in the blast furnace process energy costs are relatively high, pollution problems of associated equipment are quite severe, and capital investment requirements are often prohibitively expensive. In comparison to the blast furnace method, direct reduction permits a wider choice of fuels, is environmentally clean, and requires a much lower capital investment.

Commercial production of DRI began in the 1950s, but did not achieve significant growth until the 1970s. In 1993 world production of DRI was 23.9 million metric tons, and is expected to reach 35 million metric tons annually by the year 2000. The driving force behind this rapid increase in production is the demand for DRI as a high purity supplement to ferrous scrap in electric arc furnace steelmaking.

Physical Properties

DRI can be produced in pellet, lump, or briquette form. When produced in pellets or lumps, DRI retains the shape and form of the iron oxide material fed to the DR process. The removal of oxygen from the iron oxide during direct reduction leaves voids, giving the DRI a spongy appearance when viewed through a microscope. Thus, DRI in these forms tends to have lower apparent density, greater porosity, and more specific surface area than iron ore. In the hot briquetted form it is known as hot briquetted iron (HBI). Typical physical properties of DRI forms are shown in Table 1.

Table 1. Physical Characteristics of DRI^a

Parameter	Pellets lump	HBI
density, t m ³		
bulk	1.6–1.9	2.4–2.8
apparent	3.5	5.0–5.5
porosity, %	50	15
saturated water absorption, wt %	12–15	2–3
nominal size, mm	4–20	30 × 50 × 110

^a Produced in the MIDREX Direct Reduction Process.

HBI is produced by molding hot (ca 700°C) DRI into pillow-shaped briquettes using a pocketed roll press. HBI is almost twice as dense as nonbriquetted DRI and it has substantially less surface area, which makes it 100 times more resistant to reoxidation. It is stronger and more massive, making it more resistant to fines generation, and it takes up less volume for storage and shipping owing to its high bulk density. It has minimum water absorption when saturated, thus it is ideally suited for merchant applications where shipping, handling, and storage characteristics are important.

Chemical Properties

DRI retains the chemical purity of the iron ore from which it is produced, therefore it tends to be very low in residual elements such as copper, chrome, tin, nickel, and molybdenum. Typical ranges of DRI chemical compositions are shown in Table 2.

Table 2. Composition of DRI Produced in the MIDREX Process

Parameter	Pellet lump	HBI
iron, wt %		
total	90–94	90–94
metallic	83–89	83–89
metallization, %	92–95	92–95
FeO, wt %	6.5–9.1	6.5–9.1
C, wt %	1.0–2.5	0.8–1.2
gangue, wt %	2.8–6.0	2.8–6.0
P, wt %	0.005–0.09	0.005–0.09
S, wt %	0.001–0.03	0.001–0.03
other	trace	trace

Metallization is defined as the percent of total iron in the DRI which has been converted to metallic iron. For example, DRI having a total iron content of 92% and a metallic iron content of 85%, has 92.4% metallization.

$$\text{metallization (\%)} = \frac{\text{metallic Fe}}{\text{total Fe}} \times 100\%$$

Reduction is the percentage of oxygen present in the ore as iron oxide which has been removed.

$$\text{reduction (\%)} = \frac{\text{oxygen removed}}{\text{initial oxygen}} \times 100\%$$

Assuming that the initial iron oxide is hematite, Fe₂O₃, and this ore is completely converted to FeO, ie, no metallic iron is formed, the reduction would be 33.33%. Thus the relationship between metallization and reduction is

$$\text{metallization (\%)} = [\text{reduction (\%)} - 33.33\%] \times 1.5$$

From this relationship it can be seen that a reduction level of 95° compares with a metallization level of 92.5° . A reduction level of 33.33° or less has a metallization level of 0° . DRI normally has at least 90° reduction or 85° metallization. Processes producing solid, partially reduced iron, ie, <90° reduced or <85° metallized, are classified as prereduction processes. The partially reduced product, called prereduced iron, is not acceptable for steelmaking but can be used as a feed for iron smelting.

Although it is theoretically possible to convert all of the iron oxide in iron ore to metallic iron, it is not economically feasible. The reduction reaction slows significantly in the last stages and to complete the reduction process would result in low production rates. In practice, it is advantageous to retain a small amount of iron oxide in the DRI. During melting in an electric arc furnace, the iron oxide in DRI reacts with carbon in the DRI to form metallic iron and carbon monoxide. The carbon monoxide foams the slag during steelmaking, and this improves the operation of the electric furnace.

The carbon content of DRI depends primarily on the direct reduction process used and the way the process is operated. Carbon content can be adjusted within limits by operating changes within the DR process. Most steelmakers prefer slightly more carbon than is required to balance the remaining FeO in the DRI. DRI from gas-based processes typically contains 1 to 2.5° carbon, mostly in the form of cementite [12169-32-3], Fe₃C. DRI containing approximately 6 to 7° carbon in the form of cementite is called iron carbide. DRI from coal-based, rotary-kiln processes contains very low (ca 0.5°) levels of carbon.

The gangue content of DRI is typically comprised of oxides such as SiO₂, Al₂O₃, CaO, MgO, TiO₂, K₂O, Na₂O, MnO, etc, and is dictated by the chemistry of the iron ore used. The phosphorus in DRI is normally in the form of P₂O₅. Sulfur content in the DRI depends on the sulfur level in the ore and reductant, and the amount of sulfur released or absorbed by the DRI during the reduction process.

Production

The reduction of iron ore is accomplished by a series of reactions that are the same as those occurring in the blast furnace stack. These include reduction by CO, H₂, and, in some cases solid carbon, through successive oxidation states to metallic iron, ie, hematite [1309-37-1], Fe₂O₃, is reduced to magnetite [1309-38-2], Fe₃O₄, which is in turn reduced to wustite [17125-56-3], FeO, and then to metallic iron, Fe. The typical reactions follow.

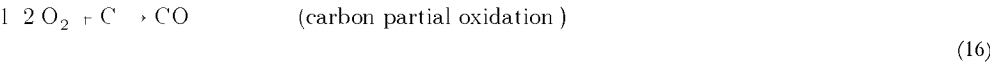
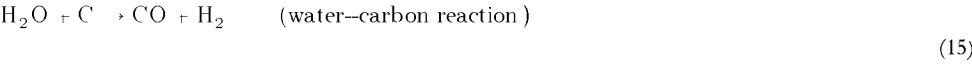
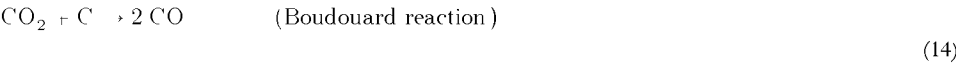
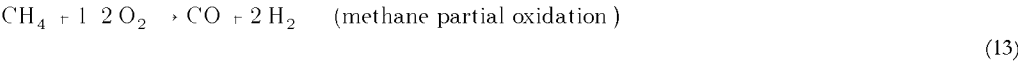
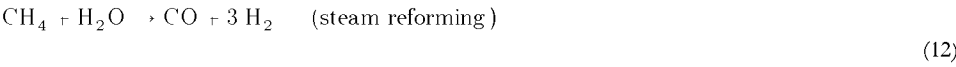
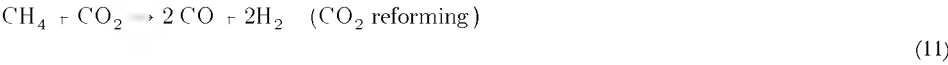
Reduction reactions



Carburization reactions



Reforming reactions



At reduction temperatures below about 1000°C, the reducing agents usually are restricted to CO and H₂. Above about 1000°C, solid carbon can

react with CO₂ and H₂O to renew the reducing potential of the gas. Above 1200°C, the metallic iron that has formed absorbs any carbon that is present, which results in melting point depression (from 1530°C) and subsequent fusing or melting of the solid.

Processes operating in the range of 1300–1530°C produce molten iron, called hot metal or pig iron. These processes are classified as direct smelting processes. Processes operating above 1530°C produce molten steel (qv) and are called direct steelmaking processes.

The equilibrium for the reactions of CO and H₂ with the oxides of iron are well established (see IRON). There is nearly complete conversion of CO to CO₂ and H₂ to H₂O for the reduction of Fe₂O₃ to Fe₃O₄. Below 570°C, Fe₃O₄ is reduced directly to Fe by CO and H₂; above 570°C, Fe₃O₄ first is reduced to FeO which then is reduced to Fe. In the reduction of Fe₃O₄ to FeO, the conversions of CO and H₂ increase with increasing reaction temperature. However, in the reduction of FeO to Fe, the conversion for H₂ increases with increasing reaction temperature, whereas that for CO decreases. This decrease of equilibrium CO conversion with increasing temperature for the reduction of FeO to Fe is not a limitation on the overall conversion because most DR processes are operated using countercurrent flow of solids and reducing gases. Thus the spent reducing gas leaves in contact with the entering solids which are in their highest oxidation state, and the equilibrium for the reduction of Fe₃O₄ to FeO governs the final gas composition. For DR processes that are based on reduction using mixtures of CO and H₂, the final gas composition usually satisfies the equilibrium for the water gas reaction at the exit temperature, ie,

$$\text{CO}_2 + \text{H}_2 \rightleftharpoons \text{CO} + \text{H}_2\text{O}, \quad K_{\text{eq}} = \frac{[\text{CO}][\text{H}_2\text{O}]}{[\text{CO}_2][\text{H}_2]}$$

(18)

The description given applies to DR processes that are based on the use of gaseous reductants in shaft furnaces, batch retorts, and fluidized beds. In the processes that use solid reductants, eg, coal (qv), the reduction is accomplished to a minor extent first by volatiles and reducing gases that are released as the coal is heated and then by CO that is formed by gasification of fixed carbon contained in the coal char with CO₂. Reduction by solid carbon and coal volatiles in kilns is insignificant.

The energy requirements for DR processes are related directly to the heats of reaction for the reduction reactions over the temperature range of practical interest for DR processes. A summary is presented in Table 3.

Table 3. Heats of Reaction for Reductions in DR Processes, kJ/kg Fe^{a,c}

Temperature, °C	Fe ₂ O ₃ —Fe ₃ O ₄			Fe ₃ O ₄ —Fe _{0.95} O			Fe _{0.95} —Fe		
	H ₂	CO	C	H ₂	CO	C	H ₂	CO	C
600	800	6,873	21,77	17,38	7,27	54,945	16,15	22,328	159,164
			1	0	1		4		
700	1,829	7,744	20,77	16,25	6,41	53,886	15,47	22,006	158,716
			9	8	7		1		
800	2,373	8,129	20,25	15,88	6,30	53,539	16,45	20,026	159,830
			6	2	0		9		
900	2,541	8,142	20,09	15,61	6,29	53,284	16,16	19,331	159,574
			3	4	2		6		
1,000	2,733	8,183	19,89	15,46	6,39	53,112	15,67	18,858	159,018
			2	3	2		2		
1,100	2,951	8,251	19,64	15,41	6,59	53,032	14,21	19,364	157,410
			9	7	7		6		
1,200	3,194	8,347	19,36	15,48	6,91	53,037	12,82	19,821	155,782
			9	4	1		2		
1,300	3,462	8,468	19,05	15,66	7,33	53,129	11,49	20,231	154,133
			0	4	0		9		

^a All heats of reaction are based on stoichiometric equation for 1 mol Fe in reactant oxide.
^b Conventional signs are employed: negative values are exothermic; positive values are endothermic.
^c To convert kJ/kg Fe to Btu/lb Fe, multiply by 0.43.

The reduction of Fe₂O₃ to Fe₃O₄ using H₂ and CO is mildly exothermic and the reduction of FeO to Fe using CO is moderately exothermic. The other reactions are moderately endothermic, with the exception of the reduction of Fe₃O₄ to FeO and of FeO to Fe by carbon; both are highly endothermic. The heat of reduction is not a principal factor in establishing the energy requirements for the gas-based DR processes. Instead, the energy requirements for these processes come mainly from the energy that is needed to generate the reducing gas, which in most cases is catalytic steam (qv) or CO₂ reforming of natural gas, and the chemical energy in the reductants that are consumed in the process (see GAS, NATURAL). For the DR processes that are conducted in shaft furnaces, total energy requirements can be minimized by recovering most of the sensible heat in the DRI and in the spent reducing gases. Such recovery is not practical in the fluidized-bed processes. For the DR processes based on the direct use of solid reductants, the highly endothermic reaction of CO₂ and solid carbon is compensated for by the highly exothermic combustion reactions in the free space above the reaction bed.

The productivity of DR processes depends on chemical kinetics, as well as mass and heat transport factors that combine to establish the overall rate and extent of reduction of the charged ore. The rates of the reduction reactions are a function of the temperature and pressure in the reduction beds, the porosity and size distribution of the ore, the composition of the reducing gases, and the effectiveness of gas–solid contact in the reduction beds. The reduction rate generally increases with increasing temperature and pressure up to about 507 kPa (5 atm).

Reduction of iron ore containing magnetite in gas-based DR processes is difficult owing to the massive structure of magnetite which hinders gaseous diffusion. If the magnetite is first oxidized to hematite, the reduction proceeds much more rapidly, because of a physical change in the crystal structure which opens up the structure and enhances gaseous diffusion.

Direct Reduction Processes

In 1993, 23.9 million metric tons of DRI were produced worldwide. Five principal processes produced 95.6% of this total. Natural gas-based direct reduction accounts for 92.5% of worldwide production and coal-based direct reduction accounts for the other 7.5%. A comparison of the five principal processes is given in Table 4.

Table 4. Comparison of Direct Reduction Processes

Parameter	MIDREX	HYL I	HYL III	SL/RN	FIOR
world production ^a , 1993, t × 10 ³ (%)	16.0 (67.1)	3.4 (14.2)	2.3 (9.8)	0.7 (3.0)	0.4 (1.5)

reduction vessel	shaft	batch retort	shaft	rotary kiln	fluid bed
reductant source	natural gas	natural gas	natural gas	coal	natural gas
iron oxide form	pellet lump	pellet lump	pellet lump	pellet lump	sized fines
DRI					
form	pellet lump or HBI	pellet lump	pellet lump or HBI	pellet lump	HBI
metallization, %	92–95	85–90	90–93	92–93	92–93
carbon content, %	1–2.5	2–2.5	1.5–3	0.2–0.5	1–1.5
reduction					
temperature, °C	760–900	870–1030	850–925	1000–1100	690–780
pressure, kPa	30	253–456	507	0	1115
time, h	4–6	6–9	4–8	8–10	6–7
reformer type	catalytic H ₂ O + CO ₂	catalytic steam	catalytic steam	none	catalytic steam
reducing gas H ₂ / CO ratio	1.5–1.8	5.35	2.25–4.65		9–10
reducing gas (H ₂ + CO) / (H ₂ O + CO ₂)	11–12	11	4.33–19		12–14
specific productivity ^d , t / (dm ³ · h)	7–12	1–2	4–12	0.3–0.5	2–3
range unit capacity installed, t × 10 ⁶ / yr	0.33–1.0	0.23–0.7	0.25–0.75	0.04–0.18	0.4
consumption per ton of DRI					
iron oxide, t	1.45	1.45	1.45	1.465	1.6–2.0
natural gas, GJ ^e	10	15–21	10–11		15–27
coal, t				0.8 ^f	
electricity, kWh	100–130	100	0–100	60–80	250
water, m ³	1.5	2.5	1.8	2–3	2.5

^a Ref. 1.
^b Other processes accounted for 1.1 × 10⁶ t or 4.4% of total 1993 world production.
^c To convert kPa to psig, multiply by 0.145.
^d Based on reactor volume.
^e To convert GJ to Btu, multiply by 1.054 × 10¹².
^f Dolomite [17069-72-6] is also added at a rate of 57 kg / t of DRI.

MIDREX Process. The primary components of a MIDREX process plant include the shaft furnace, reformer, and heat recuperator. These components are supported by ancillary systems for handling iron ore, gas, water, and direct reduced iron. A flow sheet is shown in Figure 1.

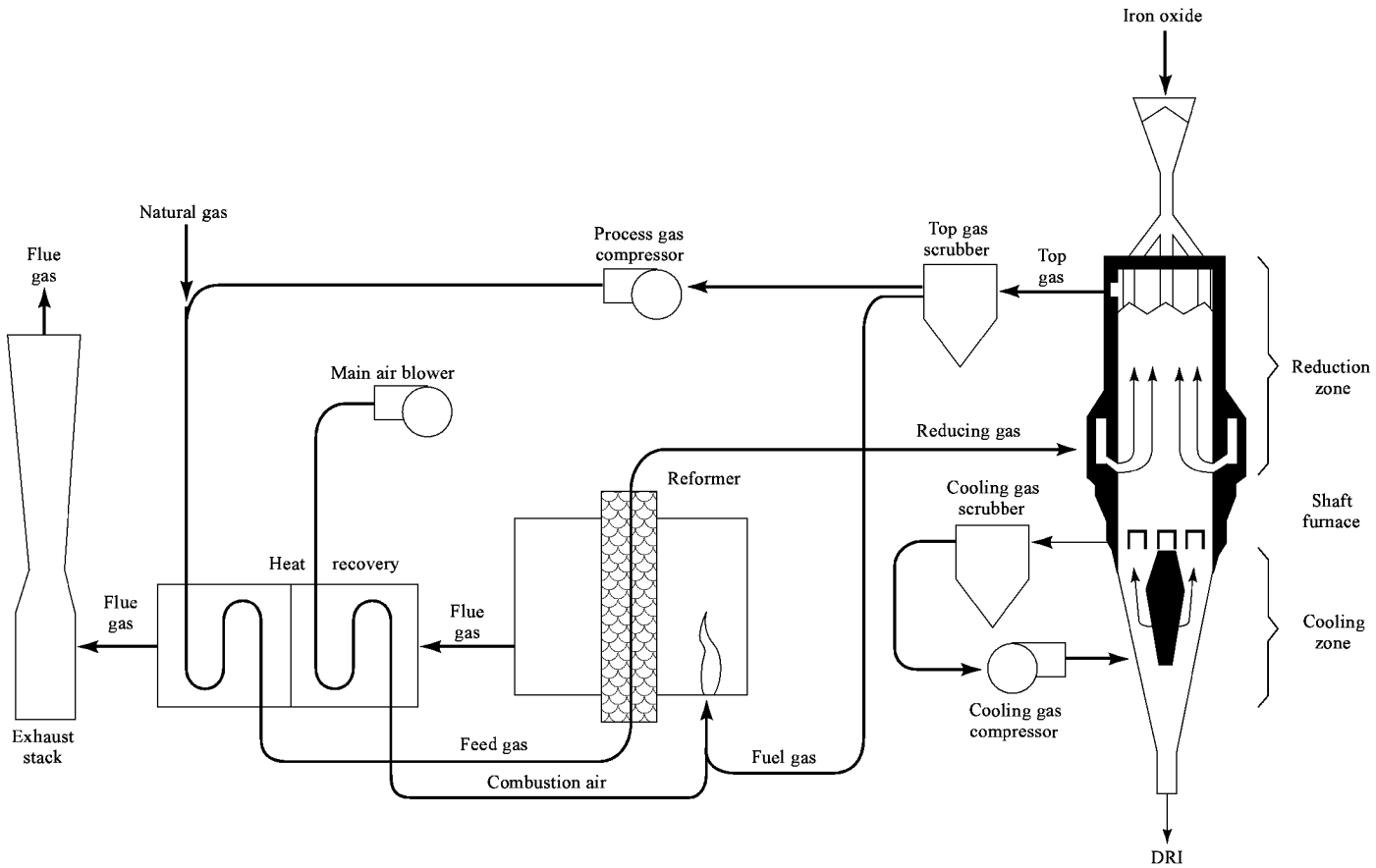


Fig. 1. The MIDREX process.

Reduction is carried out continuously in the shaft furnace. Iron oxide in pellet (6–16 mm) or lump (10–35 mm) form is fed to the top of the furnace, flows downward by gravity, and is discharged from the bottom in the form of DRI. The shaft furnace has two separate processing zones, both of which use recycled gas. In the upper reduction zone, iron oxide is preheated and reduced by counterflowing reducing gas containing hydrogen and carbon monoxide. In the lower cooling zone, the reduced product is carburized and cooled by counterflowing cooling gas.

When incorporating hot briquetting in the MIDREX process, the cooling gas circuit is eliminated, and the hot DRI is continuously discharged from the shaft furnace into a hopper and directly fed into a hot briquetting machine. The resulting HBI is continuously discharged from the hot briquetting machine, separated into individual briquettes, and cooled.

Reducing gas is generated at low pressure in the reformer by catalytically reforming a preheated mixture of fresh natural gas and recycled top gas from the shaft furnace. The reformer is a gastight, refractory-lined furnace containing alloy tubes filled with catalyst. The preheated gas mixture flows through the catalyst bed where it is heated and reformed. The hot (1000°C) reducing gas leaves the reformer at near equilibrium condition, containing 90 to

92% H₂ plus CO. Therefore, the reducing gas can be fed directly to the shaft furnace.

HYL I Process. In the HYL I process (Fig. 2), four batch retorts are operated sequentially through four steps to achieve a semicontinuous operation. While one retort is discharging DRI and being filled with pellets, two other retorts are reducing pellets and a fourth retort is carburizing and cooling DRI. Reducing gas is generated in a steam reformer and must be quenched to remove excess water vapor. Heat is transferred to the cold reducing gas by using it to carburize and cool the DRI in one of the retorts. The gas is quenched after passing through each retort and is reheated by a combination of indirect firing and partial oxidation of residual hydrocarbons to about 1050°C prior to the next stage of reduction. Reducing gas flows downward through the static bed inside the retorts and leaves a reduction gradient through the bed. The product is typically 85 to 90% metallized with 2.0 to 2.5% carbon. An auger is available if necessary to help remove DRI from the retorts.

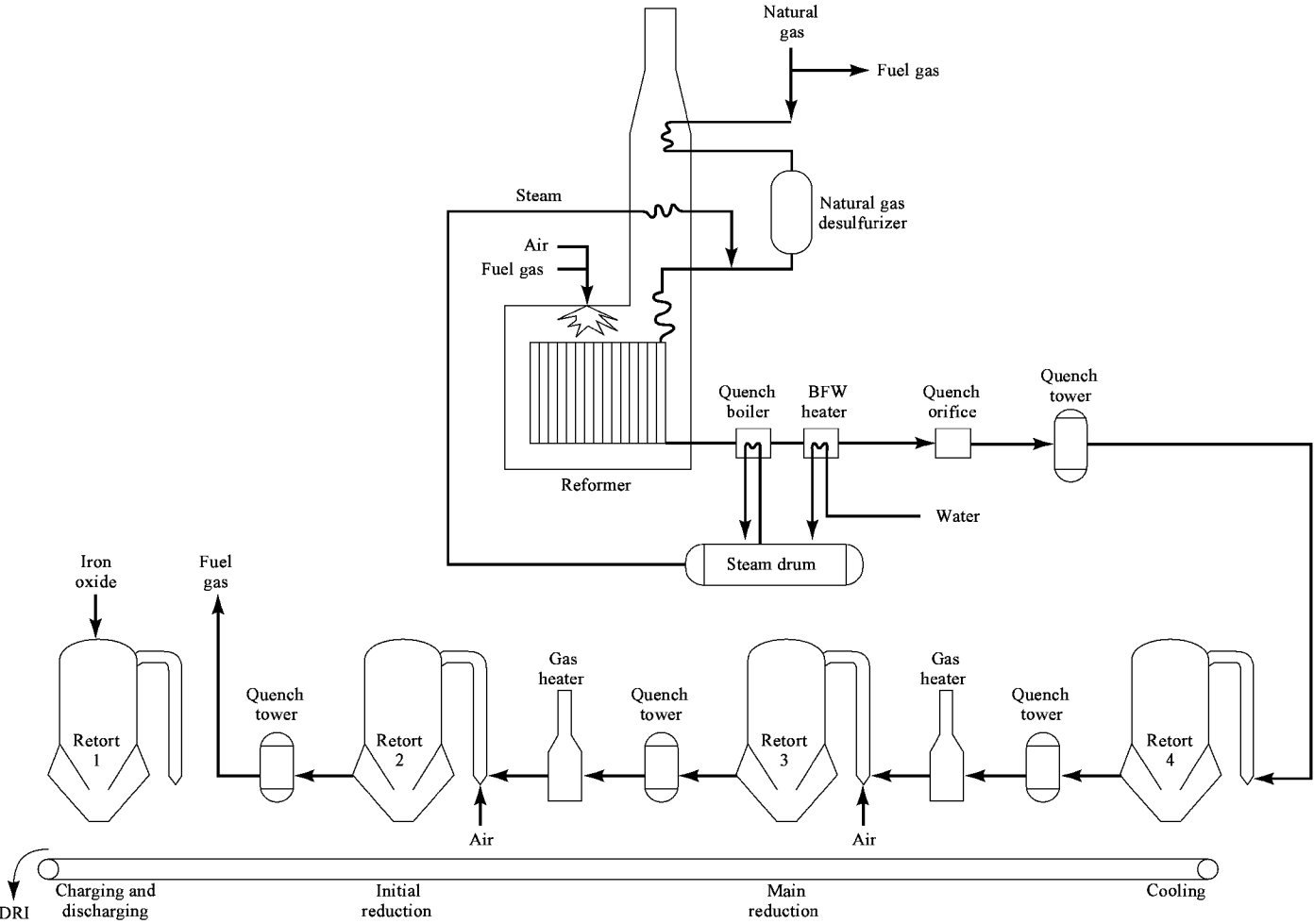


Fig. 2. The HYL I process, where BFW = boiling feed water.

HYL III Process. The HYL III process is similar to the MIDREX process, however, it uses a conventional steam reformer and pressurized shaft furnace. As shown in Figure 3, sized iron ore (pellet or lump) is charged via lock hoppers into a pressurized shaft furnace wherein the ore is heated, reduced, carburized, and cooled as it descends by gravity. The upper reduction zone of the shaft furnace is separated from the lower cooling zone by an isobaric zone. The cooled product is discharged via a rotary valve and lock hoppers onto a conveyor belt. In the case of hot briquetting, the cooling gas circuit is eliminated and the hot DRI is discharged through lock hoppers into the hot briquetting units.

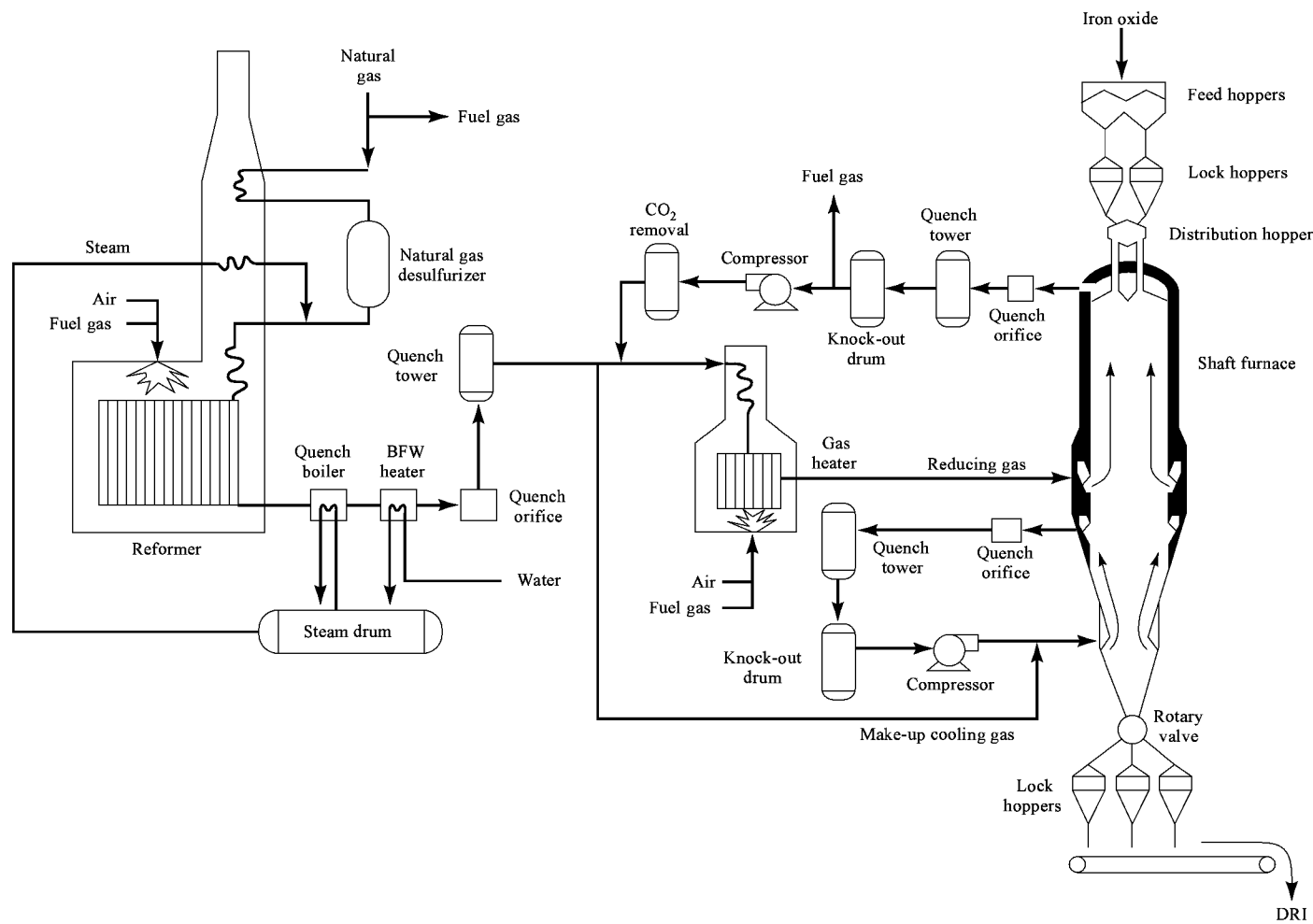


Fig. 3. The HYL III process, where BFW = boiling feed water.

Fresh reducing gas is generated by reforming natural gas with steam. The natural gas is heated in a recuperator, desulfurized to less than 1 ppm sulfur, mixed with superheated steam, further preheated to 620°C in another recuperator, then reformed in alloy tubes filled with nickel-based catalyst at a temperature of 830°C. The reformed gas is quenched to remove water vapor, mixed with clean recycled top gas from the shaft furnace, reheated to 925°C in an indirect fired heater, and injected into the shaft furnace. For high (above 92%) metallization a CO₂ removal unit is added in the top gas recycle line in order to upgrade the quality of the recycled top gas and reducing gas.

SL/RN Process. In the SL/RN process (Fig. 4), sized iron ore, coal, and dolomite are fed to the rotary kiln wherein the coal is gasified and the iron ore is reduced. The endothermic heat of reduction and the sensible energy that is required to heat the reactants is provided by combustion of volatiles and carbon monoxide leaving the bed with air introduced into the free space above the bed. The temperature profile in the kiln is controlled by radial air ports in the preheat zone and axial air ports in the reduction zone. Part of the coal is injected through the centerline of the kiln at the discharge end. The hot reduced iron and char is discharged into an indirect rotary drum cooler. The cooled product is screened and magnetically separated to remove char and ash.

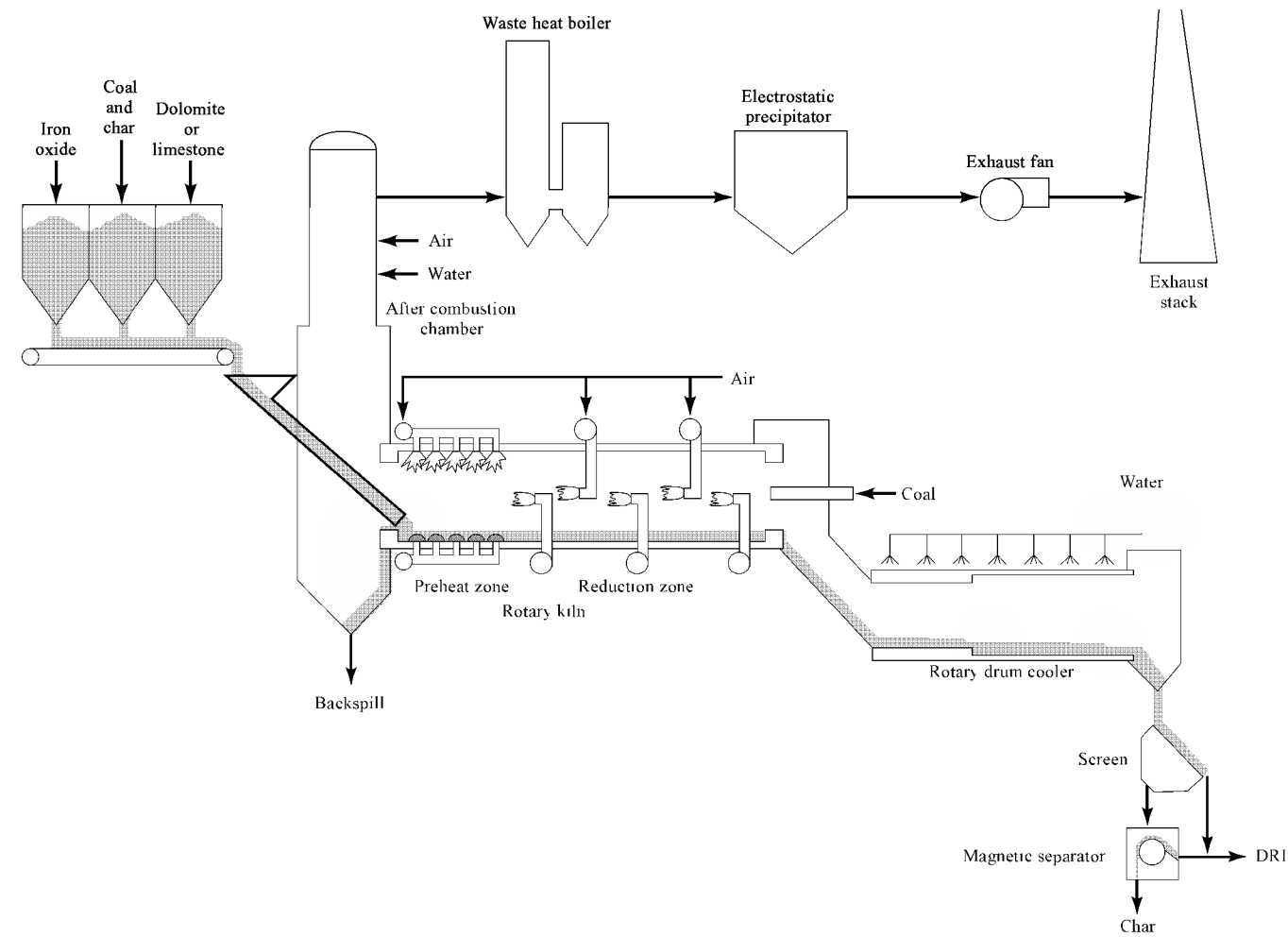


Fig. 4. The SL RN process.

The latest installations incorporate a waste heat boiler in the off-gas cleaning system to recover sensible heat from the rotary kiln off-gas. There is sufficient sensible heat in the off-gas from the SL RN process to generate 500 to 700 kWh t of DRI, depending on the type of reductant used.

FIOR Process. In the FIOR process, shown in Figure 5, sized iron ore fines (0.04–12 mm) are dried in a gas-fired rotary dryer. A skip hoist delivers the dry fines to lock hoppers for pressurizing. The fines pass through four fluidized-bed reactors in series. Reactor 1 preheats the ore to 760°C in a nonreducing atmosphere. Reactors 2, 3, and 4 reduce the ore at 690–780°C. At higher (ca 810°C) temperatures there is a tendency for the beds to defluidize as a result of sticking or bogging of the reduced material.

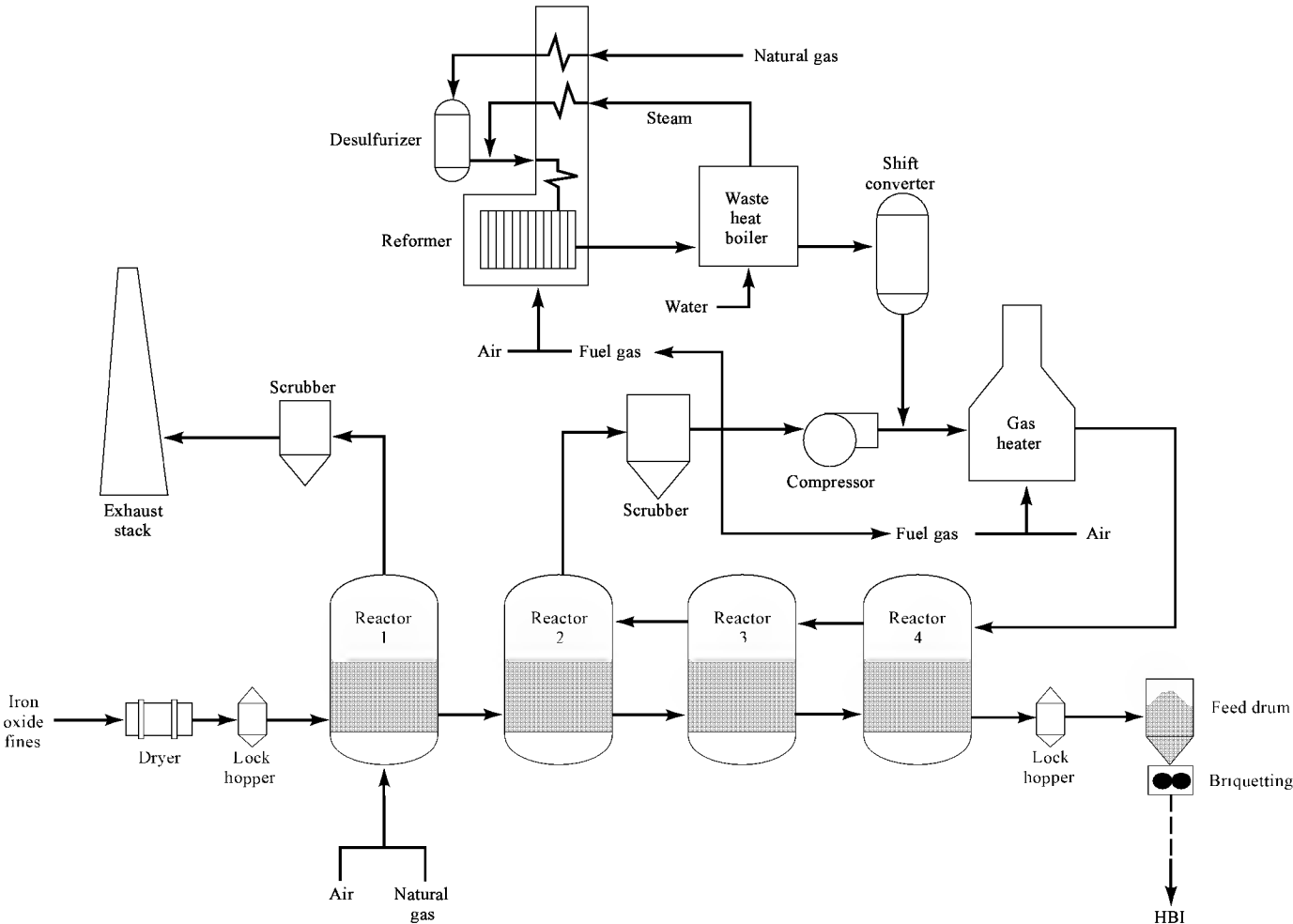


Fig. 5. The FIOR process.

The hot reduced fines are pneumatically transported to an atmospheric pressure holding drum from which they are fed to the briquetting machines. The hot briquettes are separated, cooled on a circular grate, and delivered to an outdoor storage pile.

Reducing gas is generated from natural gas in a conventional steam reformer. The natural gas is preheated, desulfurized, mixed with steam, further heated, and reformed in catalyst-filled reformer tubes at 760°C. The reformed gas is cooled to 350°C in a waste heat boiler, passed through a shift converter to increase the H₂ content, mixed with clean recycled top gas, heated to 830°C in an indirect-fired heater, then injected into reactor 4.

The reducing gas is distributed in reactor 4 by an alloy grid, passes through the fluid bed, then exits the reactor via cyclones. The gas passes through reactors 3 and 2 so that a counter flow between gas and solids is established. The spent reducing gas is scrubbed to remove dust and water vapor. Part of the cleaned top gas is recycled and the remainder is used as fuel.

Other DR Processes. The other DR processes, eg, the CODIR, DRC, ACCAR, and Dav Steel processes, make up 4.4% of worldwide production and mostly consist of coal-based, rotary-kiln processes. All of these are similar to the SL RN process. In addition, one small coal-based, shaft-furnace plant based on the Kinglor-Metor process is operating.

DR Processes Under Development. The 1990s have seen continuous evolution of direct reduction technology. Short-term development work is focusing on direct reduction processes that can use lower cost iron oxide fines as a feed material. Use of fines can represent a \$20–30 t (20%) savings in DRI production cost compared to use of pellets or lump ore. Some examples of these processes include FASTMET, Iron Carbide, CIRCOFER, and an improved version of the FIOR process.

In the FASTMET process iron oxide fines (minus 0.1 mm), pulverized coal, and binder are mixed together and pelletized. The green pellets are heated in a dryer to remove moisture and fed to a rotary hearth furnace, where the pellets are placed on a flat rotating surface (hearth) in an even layer one to two pellets deep. As the hearth rotates the pellets are heated to 1250–1350°C, and the iron oxide is reduced to metallic iron in 6 to 10 minutes.

The iron carbide process is a low temperature, gas-based, fluidized-bed process. Sized iron oxide fines (0.1–1.0 mm) are preheated in cyclones or a rotary kiln to 500°C and reduced to iron carbide in a single-stage, fluidized-bed reactor system at about 590°C in a process gas consisting primarily of methane, hydrogen, and some carbon monoxide. Reduction time is up to 18 hours owing to the low reduction temperature and slow rate of carburization. The product has the consistency of sand, is very brittle, and contains approximately 6% carbon, mostly in the form of Fe₃C.

The CIRCOFER process uses a two-stage fluidized-bed reactor system to gasify coal and reduce iron oxide fines. The coal gasification and initial reduction is performed in a circulating fluidized bed and final reduction occurs in a conventional fluidized bed. The spent reducing gases are recycled after removing water vapor and carbon dioxide. The metallized product is magnetically separated at 700°C prior to hot briquetting.

Long-term development work is focusing on direct smelting technologies. Multimillion dollar development programs are underway in the United States (AISI Direct Ironmaking process), Australia (Hismelt process), and Japan (DIOS process). A direct smelting process, called the COREX process, already is in commercial operation in South Africa.

Handling, Shipping, and Storing

In handling, shipping, and storing DRI, care should be taken to avoid oxidation. Millions of tons of DRI in pellet and lump form have been shipped by barge, ocean vessel, truck, and rail. The key to avoiding oxidation is simply to keep the material cool and dry. The chemical reactions involved have been well documented. In general, oxidation of DRI takes place in two forms: reoxidation and corrosion (2).

Reoxidation occurs when the metallic iron in hot DRI reacts with oxygen in the air to form either Fe₃O₄ or Fe₂O₃. The reaction continues as long as the DRI remains hot and sufficient oxygen is available. Because reoxidation reactions are exothermic and DRI is a good insulator, it is possible that once reoxidation begins inside a pile, the DRI temperature increases and accelerates the reoxidation rate. Although the inner core of the pile may reach temperatures up to the fusion point of iron, the maximum temperature of the outer parts of the pile will be much lower because of heat dissipation.

Corrosion occurs when the metallic iron in DRI is wetted with fresh or salt water and reacts with oxygen from air to form rust, Fe(OH)₃. The corrosion reactions continue as long as water is present. Because water evaporates at approximately 100°C, corrosion reactions have a low temperature limit even though the reactions are exothermic. Small amounts of hydrogen may be generated when DRI reacts with water. However, this poses no safety problem as long as proper ventilation is provided.

Allowing DRI to become wet does not necessarily cause it to overheat. When large piles of DRI are wetted with rain, the corrosion reactions are limited to the outer surface area of the pile and the resultant heat from the corrosion reactions is dissipated into the atmosphere. However, if water penetrates into the pile from the bottom, or if wet DRI is covered with dry DRI, the heat from corrosion reactions can build up inside the pile to the point where rapid reoxidation begins. Corrosion occurs significantly faster with salt water than with fresh water. DRI saturated with water can cause steam explosions if it is batch charged into an electric arc furnace.

In comparison, HBI is almost twice as dense as DRI, and thus does not absorb as much water and is much more resistant to reoxidation and corrosion. Several methods of passivating DRI to make it more resistant to reoxidation and corrosion have been developed, but none has been as effective as hot briquetting. Guidelines for offshore shipping of pellet lump DRI and HBI have been prepared by the International Maritime Organization.

Economic Aspects

The demand for DRI varies depending on local market conditions. In industrialized countries, DRI primarily is used as a supplement to scrap for controlling residual elements in electric arc furnace steelmaking. In regions where scrap is scarce, DRI is used as a replacement in production of all grades of steel. In 1993, Latin America produced 9.4 × 10⁶ t (39.3%) of the world's DRI. Middle East North Africa produced 6.1 × 10⁶ t (25.6%), Asia Oceania produced 4.4 × 10⁶ t (18.4%), and CIS Eastern Europe produced 1.7 × 10⁶ t (7.1%). North America produced 1.2 × 10⁶ t (5.0%); Africa, 0.9 × 10⁶ t (3.8%); and Western Europe, 0.2 × 10⁶ t (0.8%) (1). Nearly 79% of the DRI produced is consumed in steel mills adjacent to the DR plants called captive plants. Plants which are designed to sell and ship DRI on the open market are called merchant plants.

Worldwide production of DRI has increased steadily since 1970, when 7.3 × 10⁶ t were produced, except for a slight dip in 1982 owing to a steel industry recession. By 1985 production had risen to 11.16 × 10⁶ t and in 1990 to 17.89 × 10⁶ t. Projections indicate production of 35 × 10⁶ t of DRI in the year 2000 (1).

Total merchant shipments of DRI and HBI in 1993 reached 5.1 × 10⁶ t. The primary DRI exporting countries were Venezuela, Russia, Malaysia, Trinidad, and India. The price of merchant HBI in 1993 was in the range of \$125 to \$167 t on a delivered basis. Although there are expectations that the value of merchant DRI should some day stand on its own, the historic price has been tied to the price of ferrous scrap. A general rule of thumb has been that the value of merchant DRI is comparable to prime scrap (No. 1 Bundles or No. 1 Bushelings) in industrial countries, and comparable to imported shredded scrap in developing countries (see RECYCLING, FERROUS METALS).

Uses

Over 95% of the world's DRI production is consumed in electric arc furnace steelmaking. The remaining 5% is split among blast furnaces, oxygen steelmaking, foundries, and ladle metallurgy (qv) facilities.

The primary use of DRI is as a clean supplement or replacement for the ferrous scrap charge in high quality-oriented electric arc furnace (EAF) steelmaking. By controlling the level of residual elements in the charge, steelmakers can upgrade their product mix and reduce off-grade heats. Also, a low level of residual elements in carbon steel changes its physical properties for the better.

The desired portion of DRI used in the charge depends on economics, the type of steel being produced, and the available scrap quality. When DRI represents over 35% of the charge, it is preferable to feed it continuously through the roof of the EAF. Continuous charging can eliminate the need for back charging, thus improving productivity and energy efficiency. Most EAF steelmakers purchasing DRI on the open market use 10 to 30% DRI in the charge and do not have continuous charging systems installed in the melt shops. In this case, the DRI is batch charged along with scrap using existing scrap handling equipment and practices.

DRI, in pellet lump or HBI form, can be added to the blast furnace burden to increase furnace productivity and reduce coke requirements. It can be used for short-term increases in blast furnace output when a facility is short of hot metal during times of high steel demand, or when one of several blast furnaces is down for a reline. It also can be justified if the increased output is sufficient to allow operation of fewer blast furnaces long-term.

HBI is used as a trim coolant or scrap replacement in oxygen steelmaking. In the oxygen steelmaking process, the molten steel often is overheated. Trim coolant is fed to the furnace to cool the molten steel to the desired temperature. HBI is preferred for this application because its high density ensures an effective slag penetration and complete melting in the molten steel bath. Steel yield is increased when HBI is used as a trim coolant instead of iron ore. Also, the violent reactions that can occur when using iron ore are eliminated. The relative cooling effect of various materials are as follows: scrap 1.0, HBI 1.2, and iron ore 2.0–3.0.

HBI is an effective trim coolant for molten steel in ladle metallurgy facilities, ladle refiners, ladle furnaces, and vacuum degassers. It provides cold iron units in an ideal size and density for penetrating the ladle slag and cooling the metal.

HBI has been successfully melted in cupolas (hot or cold blast), induction furnaces (coreless or channel), and electric arc furnaces. It can be a valuable charge material for ductile and malleable irons as well as steel. It is of particular value in making ductile iron castings because of its very low residual element content.

BIBLIOGRAPHY

"Iron by Direct Reduction" in *ECT* 2nd ed., Suppl. Vol., pp. 535–541, by D. L. McBride, United States Steel Corp.; in *ECT* 3rd ed., Vol. 13, pp. 754–763, by J. Feinman, United States Steel Corp.

1. *Direct From Midrex*, Vol. 19, No. 2, Charlotte, N.C., 1st quarter 1994, pp. 8–10.
2. J. A. Lepinski, D. C. Meissner, and T. J. Coyne, *Steel Times*, 10–19 (Jan. 1981).

General References

Direct from Midrex **19**(2), 8–10 (1994).

L. von Bogdandy and J. Engell, *The Reduction of Iron Ores, Scientific Basis and Technology*, Springer-Verlag, Berlin, 1971.

R. L. Stephenson, ed., *Direct Reduced Iron—Technology and Economics of Production and Use*, ISS AIME, Warrendale, Pa, 1980.

W. T. Lankford, Jr. and co-workers, eds., *The Making, Shaping and Treating of Steel*, AISE, Pittsburgh, Pa, 1985.

Direct Reduction of Iron Ore: A Bibliographical Survey, The Metals Society, London, 1979.

J. A. Lepinski
PT Perkasa Indobaja

IRON COMPOUNDS

Iron (qv) is the most abundant transition metal in the earth's crust and with the exception of aluminum is more abundant than any other metallic element. It is the lightest element of Group 8 (VIII B) of the Periodic Table and is the first metallic element in the Table that fails to attain an oxidation state equal to the number of electrons in the valence shell, ie, no Fe(VIII) is known. Compounds of iron are known in which the oxidation state of iron ranges from –II ($3d^0$) to VI ($3d^6$); but aqueous chemistry of iron is dominated by the ferrous (d) and ferric (d^5) states, which are designated iron(II) and iron(III), respectively, in the preferred nomenclature. Higher oxidation states of iron generally occur in compounds that contain terminal oxo ligands. Examples include iron(VI) in $[\text{FeO}_4]^{2-}$, iron(V) in $[\text{FeO}_4]^{3-}$, and iron(VI) in compound I of cytochrome P450. All of these are potent oxidizing agents. Lower oxidation states of iron occur in compounds that contain π -acceptor ligands such as phosphines, olefins, and carbon monoxide. Examples include iron(–II) in $\text{Na}_2[\text{Fe}(\text{CO})_4]$ and iron(0) in $\text{Fe}(\text{CO})_5$ and in $[\text{Fe}(\text{1,5-cyclooctadiene})_2]$. Many low valent iron compounds are pyrophoric in air (see CARBONYLS; COORDINATION COMPOUNDS).

The standard aqueous reduction potentials for iron are



Iron metal reacts readily with most nonmetals and dissolves in dilute acids to afford the iron(II) cation. Dissolution does not occur in chromic acid, concentrated nitric acids, or hydrogen peroxide (qv), H_2O_2 , because the metal is protected by formation of a passivating oxide film, which can be removed mechanically or by acids of coordinating anions such as HCl. Iron(II) is unstable with respect to oxidation to iron(III) in the presence of air and many other oxidizing agents. The rate of air oxidation is slow in acidic solution. Both the thermodynamic and kinetic stability of iron(II) decrease in basic solution. A consequence of the facile formation of iron(III) is that many iron(II) compounds are unstable and/or nonstoichiometric. Deficiency of iron results from the substitution of iron(III) for iron(II). The ligands coordinated to iron can change the standard potentials over a substantial range of E° values. The facile interconversion of iron(II) and iron(III) and the ability of the coordination environment to fine tune the redox potential of the couple is reflected in the large variety of functions that iron performs in biological systems.

Salts of iron(II) are known for almost all of the common anions. The exceptions, including NO_2^- , result from redox incompatibilities. Many of the salts are hydrates and are subject to either efflorescence or hydration. Aqueous solutions of the salts contain the pale green hexaaquairon(II) ion $[\text{15365-81-8}]$, $[\text{Fe}(\text{H}_2\text{O})_6]^{2+}$, if the anion is noncomplexing. Hydrolysis of the hexaaquairon(II) ion is insignificant. Iron(II) forms complexes with a wide variety of both hard and soft ligands, a behavior that contrasts with that of iron(III). Most iron(II) complexes are octahedral, but other coordination numbers and geometries are not unusual. Common geometries include tetrahedral in $[\text{FeCl}_4]^{2-}$ and $[\text{Fe}(\text{SR})_4]^{2-}$; square planar in porphyrin, phthalocyanine, and tetraazamacrocyclic compounds; and square pyramidal and trigonal bipyramidal in five-coordinate complexes. Seven- and eight-coordinate complexes are also known.

The iron(II) 5D free-ion ground state is split by octahedral fields or tetrahedral fields into 5T_2 and 5E states. Tetrahedral and high spin octahedral complexes have magnetic moments of about $4.6\text{--}4.8 \times 10^{-23} \text{ J/T}$ (5.0–5.2 Bohr magnetons). The octahedral complexes exhibit a single weak $d\text{--}d$ transition, which falls in the visible/near-ir region of the spectrum and is broadened by the Jahn–Teller effect. Strong field ligands can cause spin pairing to afford diamagnetic low spin octahedral complexes that have a $^1A_{1g}$ ground state. These may be intensely purple or red in color, owing to metal-to-ligand charge-transfer bands. Low spin octahedral iron(II) complexes are kinetically inert, as are the isoelectronic low spin cobalt(III) complexes. Examples of high, low, and intermediate spin ($S = 1$) complexes are known for square planar complexes.

The high charge density of the Fe^{3+} ion results in a strong preference for class A or hard donors such as F[–] and oxygen donors. Stable compounds of amines, phosphines, and sulfur donors are relatively few in number. As in the case of iron(II), compounds of iron(III) are most typically high spin octahedral, but other spin states, coordination numbers, and geometries are well represented. The high charge density of iron(III) is also responsible for its marked tendency to hydrolyze in aqueous solution (1). The undissociated hexaaquairon(III) ion $[\text{15377-81-8}]$, predominates only below pH ~2. At slightly higher pH, $[\text{Fe}(\text{H}_2\text{O})_5(\text{OH})]^{2+}$ forms and is then converted to the hydroxo-bridged dinuclear species $[(\text{H}_2\text{O})_4\text{Fe}(\mu\text{-OH})_2\text{Fe}(\text{H}_2\text{O})_4]^{4+}$. Further increase in the pH results in formation of additional polynuclear species, colloidal gels, and finally a precipitate of red-brown hydrated ferric oxide.

High spin octahedral iron(III) compounds are d^5 and have a 6A_1 ground state. All excited states have a different spin multiplicity. Consequently, $d\text{--}d$ transitions are spin and parity forbidden and many simple salts and complexes have little or no color. These bands are often obscured by the low energy tails into the visible of ligand-to-metal charge-transfer absorptions in the near-uv region. When strong color occurs, charge-transfer absorptions are usually responsible. The magnetic moments of high spin complexes are all very close to the spin-only value of $5.47 \times 10^{-23} \text{ J/T}$ ($5.9 \mu_B$) because of the absence of an orbital angular momentum contribution or coupling to excited states. Deviations from this value are often the result of antiferromagnetic interactions between two or more iron centers. Low spin complexes typically have magnetic moments of $2.04 \times 10^{-23} \text{ J/T}$ ($2.2 \mu_B$). In lower symmetry ligand fields such as square pyramidal, intermediate spin states with moments of $3.71 \times 10^{-23} \text{ J/T}$ ($4.0 \mu_B$) are possible.

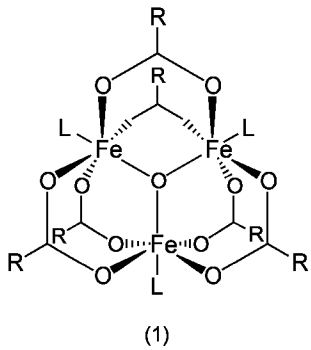
Salts and Simple Coordination Compounds

Acetates. Anhydrous iron(II) acetate $[\text{3094-87-9}]$, $\text{Fe}(\text{C}_2\text{H}_3\text{O}_2)_2$, can be prepared by dissolving iron scraps or turnings in anhydrous acetic acid (~2% acetic anhydride) under an inert atmosphere. It is a colorless compound that can be recrystallized from water to afford hydrated species. Iron(II) acetate is used in the preparation of dark shades of inks (qv) and dyes and is used as a mordant in dyeing (see DYES AND DYE INTERMEDIATES). An iron acetate salt $[\text{2140-52-5}]$ that is a mixture of indefinite proportions of iron(II) and iron(III) can be obtained by concentration of the black liquors obtained by dissolution of scrap iron in acetic acid. It is used as a catalyst of acetylation and carbonylation reactions.

Iron(III) acetate $[\text{1834-30-6}]$, $\text{Fe}(\text{C}_2\text{H}_3\text{O}_2)_3$, is prepared industrially by treatment of scrap iron with acetic acid followed by air oxidation. Iron(III) acetate is used as a catalyst in organic oxidation reactions, as a mordant, and as a starting material for the preparation of other iron-containing compounds.

Basic iron(III) acetate $[\text{10450-55-2}]$ is a brown-red material which precipitates from boiling solutions of anhydrous iron(II) or iron(III) acetate or most any soluble iron(III) salt in the presence of acetate. It is soluble in alcohols and acids but insoluble in water. The chemical formula is often given as $\text{Fe}(\text{C}_2\text{H}_3\text{O}_2)_2(\text{OH})$, but the actual composition is variable and depends on the method and conditions of preparation (2). The structure of the material is related to other basic metal acetates of the general formula $[\text{M}_3\text{O}(\text{RCO}_2)_4\text{L}_3]\text{X}$ where $L = \text{H}_2\text{O}$, CH_3OH , pyridine, or other ligand, and X is a monoanion (1). It consists of an equilateral triangle of iron atoms with a triply bridging oxygen atom at the center. Two acetate ligands bridge each edge of the triangle and a terminal ligand L coordinates to an iron at each vertex. Similar compounds exist for other carboxylates including amino acids (qv). In addition, related neutral, mixed-valent compounds of the type $[\text{Fe}^{\text{II}}\text{Fe}^{\text{III}}_2\text{O}(\text{RCO}_2)_4\text{L}_3]$ are known. The Fe_3O triangle is a structural motif in larger oxo-bridged iron

clusters, which may model the polyiron core of the iron storage protein ferritin. Basic iron acetate is used as a mordant in dyeing and printing and for the weighting of silk (qv) and felt. It is reported to affect the oxidation of saturated hydrocarbons (qv) in the presence of oxygen, acetic acid, aqueous pyridine, and zinc powder (3).



Carbonates. Iron(II) carbonate [563-71-3], FeCO₃, precipitates as a white solid when air-free solutions of alkali metal carbonates and iron(II) salts are mixed. The limited tendency of [Fe(H₂O)]²⁺ to hydrolyze is illustrated by the lack of carbon dioxide evolution in this reaction. The solid rapidly darkens when exposed to air or begins to decompose when heated to 200°C. Ferrous carbonate occurs naturally as siderite [14476-16-5] or spathic iron ore. The compound is used as a flame retardant and as an iron supplement in animal feed (see FEEDS AND FEED ADDITIVES; FLAME RETARDANTS). Ferrous carbonate redissolves in water in the presence of carbon dioxide to yield iron(II) hydrogen carbonate [6013-77-0], Fe(HCO₃)₂, which also can be formed from iron and carbon dioxide saturated water in the absence of oxygen. It undergoes air oxidation which evolves carbon dioxide and a precipitate of hydrated iron(III) oxide. This reaction accounts for the precipitation of iron from the water of many springs on exposure to air.

Citrates. Iron citrate [2338-05-8] is a compound that contains citric acid and iron(II) and iron(III) in indefinite ratios. Iron(II) citrate [23383-11-1] and iron(III) citrate [28633-45-6] are also of indefinite stoichiometry, although iron(III) citrate which contains Fe and citric acid in a 1:1 ratio [3522-50-7] is known. These compounds dissolve slowly in water and are more readily soluble in hot water. The solution chemistry of these compounds is complicated by formation of a number of monomeric and oligomeric species. All of the iron citrate compounds are used as supplements to soils and animal diets.

Iron(III) ammonium citrate [1185-57-5] is of indefinite stoichiometry. A brown hydrated compound [1332-98-5] of iron(III) ammonium citrate contains 16.5–18.5% iron, ~9% ammonia, and 65% citric acid. A green hydrated compound [1333-00-2] contains 14.5–16% iron, ~7% ammonia, and 75% citric acid. Iron ammonium citrates are water soluble but are insoluble in alcohol. The compounds are used to fortify bread, milk, and other foods (see FOOD ADDITIVES; MINERAL NUTRIENTS).

Cyanides. As a monodentate ligand, the cyanide ion coordinates to metal ions almost exclusively through the carbon atom. In this mode, it is very high in the spectrochemical series, thus most cyanide complexes are low spin (4). As a bidentate ligand the cyanide ion can bridge two metal ions by coordinating to metal ions through either the carbon or nitrogen atoms. Bonding to iron by cyanide involves synergistic σ-donation and π-acceptance by the ligand. Owing to the negative charge, the cyanide anion is a somewhat stronger donor and weaker acceptor than isoelectronic, neutral carbon monoxide (qv) (see CYANIDES).

Hexacyano Complexes. Ferrocyanide [13408-63-4] (hexakis cyanoferrate(4–)), (Fe(CN))^{4–}, is formed by reaction of iron(II) salts with excess aqueous cyanide. The reaction results in the release of 360 kJ mol (86 kcal mol) of heat. The thermodynamic stability of the anion accounts for the success of the original method of synthesis, fusing nitrogenous animal residues (blood, horn, hides, etc) with iron and potassium carbonate. Chemical or electrolytic oxidation of the complex ion affords ferricyanide [13408-62-3] (hexakis cyanoferrate(3–)), [Fe(CN)]^{3–}, which has a formation constant that is larger by a factor of 10⁷. However, hexakis cyanoferrate(3–) cannot be prepared by direct reaction of iron(III) and cyanide because significant amounts of iron(III) hydroxide also form. Hexacyanoferrate(4–) is quite inert and is nontoxic. In contrast, hexacyanoferrate(3–) is toxic because it is more labile and cyanide dissociates readily. Both complexes liberate HCN upon addition of acids.

Alkali or alkaline-earth salts of both complexes are soluble in water (except for Ba₂[Fe(CN)]) but are insoluble in alcohol. The salts of hexakis cyanoferrate(4–) are yellow and those of hexakis cyanoferrate(3–) are ruby red. A large variety of complexes arise when one or more cations of the alkali or alkaline-earth salts is replaced by a complex cation, a representative metal, or a transition metal. Many salts have commercial applications, although the majority of industrial production of iron cyanide complexes is of iron blues such as Prussian Blue, used as pigments (see PIGMENTS, INORGANIC). Many transition-metal salts of [Fe(CN)]^{4–} have characteristic colors. Addition of [Fe(CN)]^{4–} to an unknown metal salt solution has been used as a qualitative test for those transition metals.

Tetrapotassium hexakis cyanoferrate trihydrate [14459-95-1], K₄[Fe(CN)] · 3H₂O, is an efflorescent lemon yellow compound known as yellow prussiate of potash. The anhydrous material [13943-58-3] is obtained at 70°C. The compound is soluble in water and acetone, but insoluble in alcohol, ether, and ammonia. It is oxidized to hexakis cyanoferrate(3–) by oxygen in acidic solution, or by oxidants such as ozone, Cl₂, Br₂, H₂O₂, or MnO₄[–]. A large number of insoluble or slightly soluble mixed salts of the general formula K₂M^{II}[Fe(CN)] and KM^{III}[Fe(CN)] are known, eg, M^{II} = cobalt(II) [13821-10-8], copper(II) [14481-39-1], manganese(II) [15631-19-3], and nickel(II) [13601-16-6]. These have polymeric structures that contain bridging Fe–CN–M units. Many of the K₂M^{II}[Fe(CN)] compounds are useful ion-exchange (qv) materials. K₂Co[Fe(CN)] absorbs silver(I) ions from wastewater. K₂Cu[Fe(CN)] forms a semipermeable membrane which was used by van't Hoff in the measurements on which the theory of osmotic pressure was based (4).

K₄[Fe(CN)] · 3H₂O is prepared from calcium cyanide and iron(II) sulfate above 100°C. The soluble Ca₂[Fe(CN)] which forms is separated from insoluble calcium sulfate. Addition of KCl precipitates CaK₂[Fe(CN)], which is redissolved as the potassium salt by addition of potassium carbonate. Calcium carbonate is removed by filtration and K₄[Fe(CN)] · 3H₂O is crystallized by rapid cooling. K₄[Fe(CN)] · 3H₂O is used in the synthesis of other hexakis cyanoferrates(4–), in metal coatings (see METALLIC COATINGS), electroplating (qv), dyeing and printing of textiles (qv), engraving and lithography, and in the quantitative titrations of other metal salts.

Tetrasodium hexakis cyanoferrate decahydrate [14434-22-1], Na₄[Fe(CN)] · 10H₂O, or yellow prussiate of soda, forms yellow monoclinic crystals that are soluble in water but insoluble in alcohol. It is slightly efflorescent at room temperature, but the anhydrous material, tetrasodium hexakis cyanoferrate [13601-19-9], Na₄[Fe(CN)], is obtained at 100°C. The decahydrate is produced from calcium cyanide, iron(II) sulfate, and sodium carbonate in a process similar to that for the production of K₄[Fe(CN)] · 3H₂O. It is used in the manufacture of trisodium hexakis cyanoferrate, black and blue dyes, as a metal surface coating, and in photographic processing.

Tetraammonium hexakis cyanoferrate [14481-29-9], (NH₄)₄[Fe(CN)], is obtained by the addition of ammonia to [Fe(CN)]^{4–} or by the addition of ammonium sulfate to an aqueous solution of the barium or calcium salt of [Fe(CN)]^{4–}. It is soluble in water, insoluble in alcohol, and subject to air oxidation. Diammonium barium hexakis cyanoferrate [60700-20-1], (NH₄)₂Ba[Fe(CN)], and diammonium calcium hexakis cyanoferrate [60674-40-0], (NH₄)₂Ca[Fe(CN)], salts are also known.

Dibarium hexakis cyanoferrate [13821-06-2], Ba₂[Fe(CN)], is a sparingly water-soluble compound prepared by addition of a solution of

Na₄[Fe(CN)] to a concentrated solution of an appropriate barium salt. It is useful in the preparation of other [Fe(CN)]⁴⁻ salts because of the insolubility of barium sulfate.

Dicalcium hexakisicyanoferrate [13821-08-4], Ca₂[Fe(CN)], is formed as yellow crystals by reaction of liquid or gaseous HCN with iron(II) chloride in water containing Ca(OH)₂ or CaCO₃ and having pH > 8. It is used to prevent caking of other substance and serves as a useful starting material in the preparation of other [Fe(CN)]⁴⁻ salts. Examples of mixed salts include calcium dicesium hexakisicyanoferrate [15415-35-7], CaCs₂[Fe(CN)], and calcium dipotassium hexakisicyanoferrate [20219-00-5], CaK₂[Fe(CN)].

Dilead hexacyanokisferrate [14402-61-0], Pb₂[Fe(CN)], is a white precipitate that forms when lead acetate is added to Ca₂[Fe(CN)]. It is insoluble in water or dilute acids but is soluble in hot ammonium chloride or ammonium succinate solutions. It has been used as a qualitative analytical reagent in tests for cadmium and chromate.

Tripotassium hexakisicyanoferrate [13746-66-2], K₃[Fe(CN)], forms anhydrous red crystals. The crystalline material is dimorphic; both orthorhombic and monoclinic forms are known. The compound is obtained by chemical or electrolytic oxidation of hexacyanoferrate(4-). K₃[Fe(CN)] is soluble in water and acetone, but insoluble in alcohol. It is used in the manufacture of pigments, photographic papers, leather (qv), and textiles and is used as a catalyst in oxidation and polymerization reactions.

Trisodium hexakisicyanoferrate [14217-21-1], Na₃[Fe(CN)], forms red hygroscopic crystals that are soluble in water. A monohydrate [13755-37-8] and a dihydrate [36249-31-7] are also known. The sodium salt is used for many of the same purposes as the potassium salt.

Several simple salts and double salts of hexakisicyanoferrate(3-) are known. The simple salts include the ammonium [14221-48-4], (NH₄)₃[Fe(CN)], and barium [21729-04-4], Ba₃[Fe(CN)], salts. Mixed salts include (NH₄)Ag₂[Fe(CN)], [58675-53-9]; KCo[Fe(CN)], [14874-73-8]; KNi[Fe(CN)], [53295-14-0]; KCu[Fe(CN)], [53295-15-1]; and K₂Na[Fe(CN)], [31940-93-9].

All of the [Fe(CN)]⁴⁻ salts may be considered salts of ferrocyanic acid or tetrahydrogen hexakisicyanoferrate [17126-47-5], H₄[Fe(CN)], a strongly acidic, air-sensitive compound. It is soluble in water and alcohol but is insoluble in ether. It can be prepared by precipitation of an etherate by adding ether to a solution of [Fe(CN)]⁴⁻ that was acidified with concentrated sulfuric acid. Removal of the ether of solvation affords a white powder which is stable when dry but slowly turns blue in moist air because of Prussian Blue formation.

The parent acid of the hexakisicyanoferrate(3-) salts is ferricyanic acid [17126-46-4] (trihydrogen hexakisicyanoferrate). Red-brown needles are obtained by evaporation of solutions prepared by adding sulfuric acid to tribarium bis(hexakisicyanoferrate). The acid is used to prevent metal surface corrosion.

Pentacyano Complexes. Iron cyanide complexes that contain only five cyanide ligands are known as the prusside compounds. The best known of these complexes is sodium nitroprusside (sodium pentacyanonitrosylferrate(2-)), formed by the action of nitric acid or sodium nitrite on hexacyanoferrate(4-). The compound forms red, air-stable, rhombic crystals of the dihydrate [13755-38-9], Na₂[Fe(CN)₅NO]·2H₂O. It is readily soluble in water and alcohol, but the solutions are not stable. The dihydrate may be dehydrated *in vacuo* to afford an anhydrous material [14402-89-2]. Sodium nitroprusside is used as an analytical reagent for aldehydes (qv), acetone (qv), active methylene and sulfur compounds, giving an intense violet color with HS- or S²⁻ and a red color with SO² ,. It is also a useful starting material for the preparation of other prusside complexes of general formula [Fe(CN)₅L]³⁻, including L = H₂O, NH₃, NO₂, and CO. The compound is known by the trade names Nipride, Nipruss, and Nitropress and is used medically as a hypotensive (see CARDIOVASCULAR AGENTS). Blood pressure is lowered within seconds of infusion and continues only as long as administered.

Prussian Blue. Reaction of [Fe(CN)]⁴⁻ with an excess of aqueous iron(III) produces the finely divided, intensely blue precipitate Prussian Blue [14038-43-8] (tetrairon(III) tris(hexakisicyanoferrate)), Fe₄[Fe(CN)]. Prussian Blue is identical to Turnbull's Blue, the name which originally was given to the material produced by reaction of [Fe(CN)]³⁻ with excess aqueous iron(II). The solid contains or has absorbed on its surface a large and variable number of water molecules, potassium ions (if present in the reaction), and iron(III) oxide. The iron(II) centers are low spin and diamagnetic; iron(III) centers are high spin. Variations of composition and properties result from variations in reaction conditions. Rapid precipitation in the presence of potassium ion affords a colloidal suspension of Prussian Blue [25869-98-1] which has the approximate composition KFe[Fe(CN)]. Prussian Blue compounds are used as pigments in inks and paints and its formation on sensitized paper is utilized in the production of blueprints.

The structure of Prussian Blue and its analogues consists of a three-dimensional polymeric network of Fe^{II}–CN–Fe^{III} linkages. Single-crystal x-ray and neutron diffraction studies of insoluble Prussian Blue establish that the structure is based on a rock salt-like face-centered cubic (fcc) arrangement with Fe^{III} centers occupying one type of site and [Fe(CN)]⁴⁻ units randomly occupying three-quarters of the complementary sites (5). The cyanides bridge the two types of sites. The vacant [Fe(CN)]⁴⁻ sites are occupied by some of the water molecules. Other waters are zeolitic, ie, interstitial, and occupy the centers of octants of the unit cell. The structure contains three different iron coordination environments, Fe^{II}C , Fe^{III}N , and Fe^{III}N₄(H₂O), in a 3:1:3 ratio.

The intense blue color of Prussian Blue is attributed to electron transfer between the [Fe(CN)]⁴⁻ and Fe(III) ions. A related pigment called Berlin Green is obtained by oxidation of Prussian Blue. It is thought that the intense color of this other compound results only if oxidation of the [Fe(CN)] units is incomplete and some remain as hexakisicyanoferrate(4-). The compound in which only iron(III) is present, Fe[Fe(CN)] [14433-93-3], is brown and is subject to autoreduction processes.

A variety of Prussian Blue analogues of general formula (M^A)_k-[Fe(CN)]_l·xH₂O are known which have structures based on that of Prussian Blue. The occupancy of the Fe^{II}C site and the average structure of the M^A site vary with the stoichiometry of the material. There are also analogues in which iron is replaced at both sites. If M^A is a divalent metal, as in M^A = cobalt(II) [15415-49-3], copper(II) [13601-13-3], iron(II) [14460-02-7], or manganese(II) [14402-63-3], k = 2, l = 1, and the Fe^{II}C site is half occupied. The copper(II) salt is red-brown, the iron(II) salt is white when pure but is very susceptible to air oxidation to Prussian Blue, and the manganese(II) salt is green-white. For M^A = silver(I) [14308-75-6], k = 4 and l = 1. The analogues are usually insoluble in water and can be used as ion exchangers and pigments.

Formates. Iron(II) formate dihydrate [13266-73-4], Fe(HCO₂)₂·2H₂O, is a green salt which can be prepared from iron(II) sulfate and sodium formate in an inert atmosphere. The compound is slightly soluble in water and fairly resistant to air oxidation. The anhydrous salt [3047-59-4] is known.

Iron(III) formate [555-76-0], Fe(HCO₂)₃, can be obtained from iron(III) nitrate [14104-77-9] and formic acid in alcohol solution. The red compound is soluble in water but only slightly soluble in alcohol. Up to two waters of hydration may be included, in which event the color of the compound is more yellow. Aqueous solutions hydrolyze to afford basic iron(III) formates (analogous to basic acetates) and eventually a precipitate of iron hydroxide and free formate.

Fumarates. Iron(II) fumarate [141-01-5], Fe(C₄H₂O₄), is prepared by mixing hot aqueous solutions of sodium fumarate and iron(II) sulfate followed by filtration of the resulting slurry. It has limited solubility in water but is more soluble in acid solution. The compound is red-orange to red-brown and finds uses as a hematinic. A nonstoichiometric compound [7705-12-6] and iron(III) fumarate [52118-11-3], Fe₂(C₄H₂O₄)₃, are also available.

Halides. All of the anhydrous and hydrated binary halides of iron(II) and iron(III) are known with the exception of the hydrated iodide of iron(III). A large number of complex iron halides have been prepared and characterized (6).

Iron(II) fluoride [7798-28-8], FeF₂, can be prepared by the reaction of iron metal and anhydrous HF at elevated temperatures, reaction of anhydrous FeCl₂ and HF in a flow system at ca 500°C, reduction of FeF₃, or dehydration of the tetrahydrate. The solid can be sublimed and is monomeric in the gas phase above 690°C. Pure FeF₂ is a white crystalline compound and has a rutile structure in which the FeF₆⁻ octahedra are tetragonally distorted by

compression along one axis. It is sparingly soluble in water, slightly soluble in dilute HF, and insoluble in alcohol, ether, and benzene. Above 100 K, it has a magnetic moment of 5.56 BM. Iron(II) fluoride is used as a catalyst in organic fluorinations.

Iron(II) fluoride tetrahydrate [13940-89-1], FeF₂·4H₂O, is prepared by dissolving iron metal in warm hydrofluoric acid and precipitating with ethanol. The structure of the solid consists of discrete [FeF₂(H₂O)₄] octahedra in which F⁻ and H₂O are randomly distributed over the possible sites. The white solid turns brown in air and decomposes at 100°C. It is slightly soluble in water, alcohol, and ether and is soluble in dilute acid.

Iron(III) fluoride [7783-50-8], FeF₃, is prepared from FeCl₃ and anhydrous HF or other fluorinating agents in a flow system at elevated temperature. The green hexagonal crystals sublime above 1000°C. Iron(III) fluoride is slightly soluble in water, freely soluble in dilute HF, and nearly insoluble in alcohol, ether, and benzene. It is used as a catalyst in organic reactions.

Iron(III) fluoride trihydrate [15469-38-2], FeF₃·3H₂O, crystallizes from 40% HF solution in two possible crystalline forms. At low temperature the α-form, which is isostructural with α-AlF₃·3H₂O, is favored. High temperatures favor β-FeF₃·3H₂O, the structure of which consists of fluoride-bridged octahedra with one water of hydration per unit cell.

Iron(II) chloride [7758-94-3], FeCl₂, is prepared by reaction of iron and HCl at red heat, iron and a mixture of HCl and Cl₂ at 700°C in a flow system, iron and CCl₄, or by decomposition of FeCl₃ at 300°C *in vacuo*. Several methods employ reduction of FeCl₃. The compound occurs naturally as the mineral lawrencite. White, very hygroscopic crystals can be obtained by sublimation at 700°C in a stream of HCl. The compound decomposes to FeCl₃ and Fe₂O₃ on heating in air. Under normal conditions, the crystalline solid has the CdCl₂ structure. The room temperature magnetic moment is 5.44 × 10⁻²³ J/T (5.87 μ_B). The compound is soluble in water, alcohol, and acetone; slightly soluble in benzene; and insoluble in ether. It reacts with numerous ligands to form complexes. Iron(II) chloride is used as a reducing agent, as a mordant in dyeing, and in pharmaceuticals (qv) and metallurgy (qv).

Iron(II) chloride tetrahydrate [13478-10-9], FeCl₂·4H₂O, is obtained by dissolving iron metal in aqueous HCl and allowing the product to crystallize at room temperature. The solid consists of monomeric *trans*-FeCl₂(H₂O)₄ octahedra which hydrogen bond extensively with each other. Iron(II) chloride dihydrate [16399-77-2], FeCl₂·2H₂O, is obtained by crystallization at 75°C or by careful dehydration of the tetrahydrate at 105–115°C. The dihydrate consists of chains of *trans*-FeCl₄(H₂O)₂ octahedra linked by two bridging chlorides. At 150–160°C the dihydrate loses water to yield iron(II) chloride monohydrate, FeCl₂·H₂O, which at 220°C loses the remaining water of hydration.

Iron(III) chloride [7705-08-0], FeCl₃, can be prepared from iron in a stream of Cl₂ at 350°C. Pure material sublimes in the stream of Cl₂ and is obtained as hexagonal crystals which appear green to reflected light but red to transmitted light. The crystals are very hygroscopic and in moist air form a dark brown liquid which contains the hexahydrate. Iron(III) chloride is very soluble in water, alcohol, ether, and acetone, but only slightly soluble in carbon disulfide and essentially insoluble in ethyl acetate. Dissolution in water is exothermic. Iron(III) chloride melts and volatilizes at ca 300°C and boils at ca 316°C. Below 400°C, the vapor contains dimeric Fe₂Cl₆, which consists of two edge-sharing tetrahedra. At higher temperatures the vapor becomes monomeric and eventually decomposes to FeCl₂ and Cl₂. Iron(III) chloride is a convenient starting material in the syntheses of other iron(III) compounds and salts. Many adducts and substitution products are known. It is used as a chlorinating and oxidizing agent and is used in the manufacture of dyes, inks, and pigments. Iron(III) chloride is used as a coagulant in styptics and in water treatment to remove organic matter and suspended solids by producing a ferric hydroxide flocculent. It is also used to purify industrial waste gases.

Iron(III) chloride hexahydrate [10025-77-1], FeCl₃·6H₂O, is a brown-yellow to orange material that crystallizes from a solution of iron or iron salt dissolved in hydrochloric acid that contains an oxidant such as Cl₂ or nitric acid. The monoclinic crystals contain the complex salt *trans*-[FeCl₂(H₂O)₄]Cl·2H₂O. The crystals are very hygroscopic and dissolve readily in water, alcohol, acetone, and ether. In aqueous solutions, hydrolysis is extensive and iron hydroxide precipitates. A series of compounds with varying hydration can be formed by drying FeCl₃·6H₂O.

Iron(II) bromide [7789-46-0], FeBr₂, can be prepared by reaction of iron and bromine in a flow system at 200°C and purified by sublimation in nitrogen or under vacuum. Other preparative routes include the reaction of Fe₂O₃ with HBr in a flow system at 200–350°C, reaction of iron with HBr in methanol, and dehydration of hydrated forms. FeBr₂ crystallizes in a layered lattice of the CdI₂ type and has a magnetic moment of 5.30 × 10⁻²³ J/T (5.71 μ_B) at room temperature. It is air stable at 25°C, but is slowly converted to Fe₂O₃ and bromine at 310°C. The light yellow to brown hygroscopic solid is soluble in water, alcohol, ether, and acetonitrile. Iron(II) bromide forms adducts with a wide range of donor molecules. Pale green nona-, hexa-, tetra-, and dihydrate species can be crystallized from aqueous solutions at different temperatures. A hydrate of variable water content, FeBr₂·xH₂O [13463-12-2], is commercially available. Anhydrous iron(II) bromide is used as a catalyst in organic brominations and polymerization reactions.

Iron(III) bromide [10031-26-2], FeBr₃, is obtained by reaction of iron or iron(II) bromide with bromine at 170–200°C. The material is purified by sublimation in a bromine atmosphere. The structure of iron(III) bromide is analogous to that of iron(III) chloride. FeBr₃ is less stable thermally than FeCl₃, as would be expected from the observation that Br⁻ is a stronger reductant than Cl⁻. Dissociation to iron(II) bromide and bromine is complete at ca 200°C. The hygroscopic, dark red, rhombic crystals of iron(III) bromide are readily soluble in water, alcohol, ether, and acetic acid and are slightly soluble in liquid ammonia. Several hydrated species and a large number of adducts are known. Solutions of iron(III) bromide decompose to iron(II) bromide and bromine on boiling. Iron(II) bromide is used as a catalyst for the bromination of aromatic compounds.

Iron(II) iodide [7783-86-0], FeI₂, is easily prepared by direct reaction of the elements. It has a CdI₂-type structure and a magnetic moment of 5.33 × 10⁻²³ J/T (5.75 μ_B) at room temperature. The hygroscopic, red-black crystals are very soluble in water, alcohol, and ether. Aqueous solutions are readily oxidized by air. The green tetrahydrate [13492-45-0], FeI₂·2H₂O, can be obtained by evaporation at room temperature; a yellow hexahydrate can be obtained at lower temperature. The hydrated salts and their solutions turn black when heated to 50°C, but regain their original color on cooling. FeI₂ forms many adducts with donor ligands. It is used as a catalyst in organic reactions and as a source of both iron and iodine in veterinary medicine (see VETERINARY DRUGS).

Iron(III) iodide [15600-49-4], FeI₃, is prepared by the oxidative photodecarbonylation of diiodotetracarbonyliron(II) in the presence of diiodine (7). The black solid obtained is extremely hygroscopic, sparingly soluble only in dichloromethane, and decomposes to iron(II) iodide and diiodine when exposed to donor solvents such as tetrahydrofuran, acetonitrile, water, or pyridine. It also decomposes when exposed to light.

Iron halides react with halide salts to afford anionic halide complexes. Because iron(III) is a hard acid, the complexes that it forms are most stable with F⁻ and decrease in both coordination number and stability with heavier halides. No stable I⁻ complexes are known. [FeF₆(H₂O)]²⁻ is the predominant iron fluoride species in aqueous solution. The [FeF₄]⁻ ion can be prepared in fused salts. Whereas six-coordinate [FeCl₆]³⁻ is known, four-coordinate complexes are favored for chloride. Salts of tetrahedral [FeCl₄]⁻ can be isolated if large cations such as tetraphenylarsonium or tetraalkylammonium are used. [FeBr₄]⁻ is known but is thermally unstable and disproportionates to iron(II) and bromine. Complex anions of iron(II) halides are less common. [FeCl₂]²⁻ has been obtained from FeCl₂ by reaction with alkali metal chlorides in the melt or with tetraethylammonium chloride in deoxygenated ethanol.

Several complex anions are known in which one or more halide ligands are replaced by other ligands. One important example is the μ-oxo-bridged binuclear anion [Fe₂OCl]²⁻, which consists of two FeOCl₃ tetrahedra sharing the oxygen vertex. The two iron centers are strongly antiferromagnetically coupled. The anion is the first example of a μ-oxo-bridged binuclear iron(III) complex that has tetrahedral coordination and monodentate supporting ligands. The anion is finding use as a starting material for the synthesis of larger oxo-bridged iron clusters (8).

Gluconates. Iron(II) gluconate dihydrate [6047-12-7], Fe[HOCH₂(CHOH)₄CO₂]₂·2H₂O, is prepared from barium or calcium gluconate and iron(II) sulfate. It is a yellow-grown powder and has a slight odor of caramel. The compound is quite soluble in water but is nearly insoluble in alcohol. It is

used as a hematinic, in the treatment of anemia, and to color, fortify, and flavor foods (see FLAVORS AND SPICES). Isotonic solutions are available. An anhydrous salt [299-29-6] is also known. Iron(III) gluconate [38658-53-6], $\text{Fe}[\text{HOCH}_2(\text{CHOH})_4\text{CO}_2]_3$, has been examined as a nutritional supplement.

Nitrates. Iron(II) nitrate hexahydrate [14013-86-6], $\text{Fe}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$, is a green crystalline material prepared by dissolving iron in cold nitric acid that has a specific gravity of less than 1.034 g cm^{-3} . Use of denser, more concentrated acid leads to oxidation to iron(III). An alternative method of preparation is the reaction of iron(II) sulfate and barium or lead nitrate. The compound is very soluble in water. Crystallization at temperatures below 12°C affords an anhydrate. Iron(II) nitrate is a useful reagent for the synthesis of other iron-containing compounds and is used as a catalyst for reduction reactions.

Iron(III) nitrate nonahydrate [7782-61-8], $\text{Fe}(\text{NO}_3)_3 \cdot 9\text{H}_2\text{O}$, is prepared by dissolving iron in nitric acid that has a specific gravity of at least 1.115 g cm^{-3} . Acid of too high concentration passivates the iron, however. The hygroscopic, monoclinic, colorless-to-pale violet crystals are very soluble in water and soluble in alcohol and acetone. Iron(III) nitrate hexahydrate [13476-08-9], $\text{Fe}(\text{NO}_3)_3 \cdot 6\text{H}_2\text{O}$, forms colorless, cubic crystals. It is also very soluble in water. Iron(III) nitrate is used as a mordant in dyeing, weighting silks, leather tanning, as a catalyst for oxidation reactions, and as a reagent for the synthesis of other iron-containing compounds.

Oxides and Hydroxides. Iron(II) oxide [1345-25-1], FeO, is a black solid that can be obtained by heating iron in a low partial pressure of oxygen, or by heating iron(II) oxalate in a vacuum. This affords a fine, pyrophoric powder that can decompose water. Strong heating of the freshly prepared powder decreases its state of division and its reactivity. Below ca 575°C, FeO is unstable with respect to disproportionation into Fe and Fe_3O_4 , but a metastable phase can be obtained by rapidly cooling the hot product. FeO occurs naturally as the mineral wüstite [17125-56-3]. Crystalline FeO has a cubic, rock salt structure, but is always deficient in iron because of the presence of some iron(III). The solid is easily oxidized in air, is a strong base, and absorbs carbon dioxide. It is insoluble in water, alcohol, or alkali, but is readily soluble in acids. Iron(II) oxide is used in the manufacture of green, heat-absorbing glass, in ceramic mixtures, and in a variety of catalyst preparations, notably those used in ammonia synthesis and methanation.

Iron(III) oxide [1309-37-1], Fe_2O_3 , exists in two different crystalline forms. The α -form occurs naturally as the mineral hematite [1309-37-1], which is the principal ore of iron. It can be prepared synthetically by heating brown hydrous iron hydroxide oxide [20344-49-4], $\text{Fe}(\text{OH})\text{O}$, at 200°C. α - Fe_2O_3 has the corundum structure in which the oxide ions form hexagonal close-packed layers and the iron(III) ions occupy two-thirds of the octahedral sites. The γ -form occurs naturally as the mineral maghemite, and can be prepared synthetically by careful oxidation of Fe_3O_4 . γ - Fe_2O_3 has a spinel structure in which the oxide ions form cubic close-packed layers and the iron(III) ions are randomly distributed over the tetrahedral and octahedral holes. Unlike the α -form, which is paramagnetic, the γ -form is ferrimagnetic and is used as a magnetic material in the production of magnetic recording media (see MAGNETIC MATERIALS). Iron(III) oxide is insoluble in water but dissolves in hydrochloric or sulfuric acids. The color and appearance of iron(III) oxide depend on the size and shape of the particles, and the identity and amount of impurities and water present. Yellow, orange, or red pigments are known. It is used in large quantities as a red pigment for paint, rubber, ceramics (qv), and paper (qv), and in coatings for steel and other metals. Iron(III) oxide also finds use in the preparation of rare-earth iron garnets and other ferrites; as a polish agent for glass, diamonds, and precious metals (jeweler's rouge); and as a catalyst for oxidation reactions.

Triiron tetroxide [1317-61-9] (iron(II, III) oxide), Fe_3O_4 , is a mixed $\text{Fe}^{\text{II}} \text{Fe}^{\text{III}}$ oxide which occurs naturally as the mineral magnetite (lodestone). It is an important ore of iron. It can be prepared by the partial oxidation of FeO or by heating Fe_2O_3 above ca 1400°C. The black, cubic crystals have an inverse spinel structure in which the oxide ions form cubic close-packed layers, all iron(II) ions occupy octahedral interstitial sites, half of the iron(III) ions occupy octahedral sites, and the other half of the iron(III) ions occupy tetrahedral sites. The compound is strongly ferrimagnetic and has a Curie point of 860 K, at which temperature the effective magnetic moment is $3.9 \times 10^{-23} \text{ J T}^{-1}$ (4.2 μ_{B}). Iron(II,III) oxide is insoluble in water, alcohol, ether, and dilute acids but dissolves in concentrated acids. It is a fairly good conductor of electricity, owing to electron transfer between iron(II) and iron(III). Blue steel has a surface coating of iron(II,III) oxide as a corrosion-resistant film. The compound is used as a pigment for glass (qv), ceramics, and paint; in magnetic recording media; as a polishing compound; and in many catalytic preparations.

Iron(II) hydroxide [18624-44-7], $\text{Fe}(\text{OH})_2$, is prepared by precipitation of an iron(II) salt solution by strong base in the absence of air. It occurs as pale green, hexagonal crystals or a white amorphous powder. It is practically insoluble in water, fairly soluble in ammonium salt solutions, and soluble in acids and in concentrated NaOH solution. It is slowly oxidized by air. Conversion to $\text{Fe}_2\text{O}_3 \cdot x\text{H}_2\text{O}$ is eventually complete.

Iron(III) hydroxide [1309-33-7], FeHO_2 , is a red-brown amorphous material that forms when a strong base is added to a solution of an iron(III) salt. It is also known as hydrated iron(III) oxide. The fully hydrated $\text{Fe}(\text{OH})_3$ has not been isolated. The density of the material varies between 3.4–3.9 g cm^{-3} , depending on its extent of hydration. It is insoluble in water and alcohol, but redissolves in acid. Iron(III) hydroxide loses water to form Fe_2O_3 . Iron(III) hydroxide is used as an absorbent in chemical processes, as a pigment, and in abrasives. Salt-free iron(III) hydroxide can be obtained by hydrolysis of iron(III) alkoxides.

Ferrites, Garnets, and Ferrates. Iron in oxidation states +3 and higher forms numerous oxide compounds that formally appear to contain oxo ligands (O^{2-}). Many of these have interesting magnetic and chemical properties. The ferrites (FeO_2) and garnets are really mixed metal oxides (9). Ferrites (qv) of the alkali metals can be obtained by fusing iron(III) oxide with the alkali metal chloride, carbonate, or hydroxide, or by decomposing the alkali iron(VI) oxo compound in boiling water. Sodium ferrite [12062-85-0], NaFeO_2 , occurs as brown hexagonal plates or needles which are soluble in dilute HCl. Lithium ferrite [12022-46-7] is also known. Ferrites of divalent metals are prepared by heating iron(III) oxide with the carbonate of the desired metal, or by addition of strong base to a solution of the M(II) and iron(III) salts. Examples include magnesium ferrite [12068-86-9], calcium ferrite [12013-33-1], barium ferrite [12009-00-6], and zinc ferrite [1317-55-1] which occurs naturally as the mineral franklinite. Ferrite compounds are spinels and have the general formula $\text{M}^{2+}\text{Fe}^{3+}_2\text{O}_4$. Some adopt normal spinel structures, in which the M(II) ions occupy tetrahedral sites in the cubic oxide lattice and the Fe(III) ions occupy octahedral sites. Others have the inverse spinel structure in which one half of the Fe(III) ions occupy tetrahedral sites and the other half occupies octahedra sites. Inverse spinels are ferrimagnetic. One such material is iron(II,III) oxide. The inverse spinel ferrites are used in magnetic recording media, as cores in high frequency transformers, and in computer memory systems. Hexagonal ferrites, such as $\text{BaFe}_{12}\text{O}_{19}$, are used to construct permanent magnets (10). Garnets have the general formula $\text{M}_3\text{Fe}_5\text{O}_{12}$, where M is trivalent, and are useful in microwave applications (see MICROWAVE TECHNOLOGY).

Mixed oxides of Fe(IV) can be prepared by heating iron(III) oxide with a metal oxide or hydroxide in oxygen at elevated temperatures. These black compounds have general formulas M_4FeO_4 , M monovalent, or M_2FeO_4 , M divalent, but do not contain discrete $[\text{FeO}_4]^{4-}$ ions. They are readily decomposed by mineral acids to iron(III) and oxygen.

Compounds of iron(V) are extremely rare. K_3FeO_4 has been prepared by heating K_2FeO_4 with KOH in oxygen at 700–800°C. It appears to contain tetrahedral $[\text{FeO}_4]^{3-}$ anions. An impure sodium salt has also been prepared.

The best known oxoanion of iron is the ferrate(VI) prepared by oxidizing a suspension of hydrous iron(III) oxide in concentrated alkali with potassium hypochlorite or by anodic oxidation of iron in concentrated alkali. Crystals of potassium ferrate [13718-66-6], K_2FeO_4 , are deep purple, orthorhombic, and contain discrete tetrahedral $[\text{FeO}_4]^{2-}$ anions. Barium ferrate [13773-23-4] can be precipitated from solutions of soluble ferrate salts. Other ferrate salts include calcium ferrate [35764-67-1], and sodium ferrate [13773-03-0]. The magnetic moments of these materials are 2.63.0 $\times 10^{-23} \text{ J T}^{-1}$ (2.8–3.2 μ_{B}), which is consistent with the expectation of two unpaired electrons. The $[\text{FeO}_4]^{2-}$ ion is an extremely strong oxidizing agent, oxidizing NH_3 to N_2 at room temperature, and in neutral or acidic solutions rapidly oxidizing water to oxygen. The $[\text{FeO}_4]^{2-}$ is a stronger oxidant than permanganate and has found use in the oxidation of organic compounds.

Perchlorates. Iron(II) perchlorate hexahydrate [13922-23-8], $\text{Fe}(\text{ClO}_4)_2 \cdot 6\text{H}_2\text{O}$, is prepared by dissolving iron in cold, dilute perchloric acid or by dissolving FeS in perchloric acid. It crystallizes in hygroscopic, light green hexagonal prisms which are stable in dry air and extremely soluble (0.978 g mL^{-1} H_2O at 0°C) in water and alcohol. It is susceptible to air oxidation in aqueous solution and decomposes above 100°C. Yellow iron(III) perchlorate

hexahydrate [13537-24-2], $\text{Fe}(\text{ClO}_4)_3 \cdot 6\text{H}_2\text{O}$, is also extremely soluble in water (1.198 g mL H_2O at 0°C).

Sulfates. Iron(II) sulfate heptahydrate [7782-63-0], $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, forms blue-green monoclinic crystals that are very soluble in water and somewhat soluble in alcohols. It is known by many other names including cupperas, green vitriol, and iron vitriol. The compound is efflorescent in dry air. In moist air, the compound oxidizes to yellow-brown basic iron(III) sulfate. Aqueous solutions tend to oxidize. The rate of oxidation increases with an increase in pH, temperature, and light. The compound loses three waters of hydration to form iron(II) sulfate tetrahydrate [20908-72-0], $\text{FeSO}_4 \cdot 4\text{H}_2\text{O}$, at 56°C. Further warming to 65°C forms white iron sulfate monohydrate [17375-41-6], $\text{FeSO}_4 \cdot \text{H}_2\text{O}$, which is stable to 300°C. Strong heating results in decomposition with loss of sulfur dioxide. Solutions of iron(II) sulfate reduce nitrate and nitrite to nitric oxide, whereupon the highly colored $[\text{Fe}(\text{H}_2\text{O})_5(\text{NO})]^{2+}$ ion is formed. This reaction is the basis of the brown ring test for the qualitative determination of nitrate or nitrite.

Iron(II) sulfate forms double salts of formula $\text{M}_2\text{SO}_4 \cdot \text{FeSO}_4 \cdot 6\text{H}_2\text{O}$ with alkali sulfates. Iron(II) ammonium sulfate [7783-85-9] (Mohr's salt), $\text{FeSO}_4 \cdot (\text{NH}_4)_2\text{SO}_4 \cdot 6\text{H}_2\text{O}$, is used as a primary standard for iron. It is soluble in water but insoluble in alcohols. Both the solid and its solution are more stable to oxidation than iron(II) sulfate.

Most iron(II) sulfate is a by-product of the steel (qv) industry. Prior to tinning, galvanizing, electroplating, or enameling, steel surfaces are dipped in sulfuric acid for cleaning (pickling) (see METAL SURFACE TREATMENTS). The resulting pickle liquor contains ca 15% iron(II) sulfate and 2–7% acid. Scrap iron is added to reduce the acid concentration to ca 0.03%. The solution is filtered, concentrated at 70°C to a specific gravity of 1.4, and is allowed to cool to room temperature which results in crystallization of the heptahydrate. Industry produces on the order of 10^6 t yr of the heptahydrate. Because supply exceeds demand, the pickling liquor presents a serious waste disposal problem. Iron(II) sulfate, along with iron(III) sulfate, and sulfuric acid can also be produced by leaching or weathering of FeS and pyrites, which is a source of acid mine drainage. Iron(II) sulfate has a large variety of uses including production of iron oxide pigments and salts, in fertilizer, as food and feed supplements, in inks (qv) and dyes, as a reducing agent, and a polymerization catalyst.

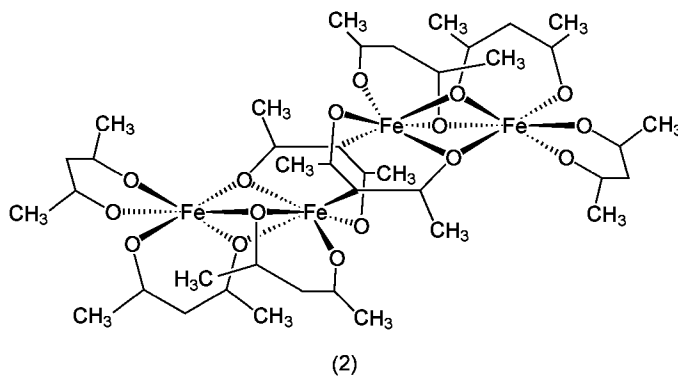
Iron(III) sulfate [10028-22-5], $\text{Fe}_2(\text{SO}_4)_3$, is a gray-white material that forms hygroscopic rhombic or rhombohedral crystals. It is slightly soluble and dissolves slowly in cold water and decomposes in hot water. Several hydrates are known including the monohydrate [43059-01-4], the hexahydrate [13761-89-2], the heptahydrate [35139-28-7], and the nonahydrate [13520-56-4]. These can be difficult to obtain pure. The commercially available hydrate [15244-10-7] contains about 20% water by weight and is yellowish in color. Alum compounds of the general formula $\text{MFe}(\text{SO}_4)_2 \cdot 12\text{H}_2\text{O}$, M monovalent, are known. Iron(III) ammonium alum dodecahydrate [7783-83-7], $(\text{NH}_4)\text{Fe}(\text{SO}_4)_2 \cdot 12\text{H}_2\text{O}$, and iron(III) potassium alum dodecahydrate [13464-29-1], $\text{KFe}(\text{SO}_4)_2 \cdot 12\text{H}_2\text{O}$, are important as mordants in dyeing. Iron(III) sulfate is obtained by oxidation of iron(II) sulfate using nitric acid or by treating iron(II) oxide with sulfuric acid. It is used in pigments, as a coagulant in water and sewage treatment, and as a mordant.

Sulfides. Three sulfides of iron are known. Iron(II) sulfide [1317-37-9], FeS, is a gray nonstoichiometric material obtained by direct reaction of iron and sulfur. The actual stoichiometry is typically $\text{Fe}_{0.9}\text{S}$. It can also be prepared by treating iron(II) solutions with alkali metal sulfide. It is found in nature as the mineral pyrrhotite [1310-50-5], which usually contains nickel as well. FeS has a NiAs structure. It is almost insoluble in water, oxidizes readily in air, and dissolves in aqueous acids with the evolution of H_2S . The above reactions represent a reasonable route for the synthesis of H_2S gas. Iron disulfide [12068-85-8], FeS_2 , can be prepared by heating Fe_2O_3 in H_2S . FeS_2 is found in nature as the minerals pyrite [1309-36-0] (fool's gold) and marcasite [1317-66-4], both of which have a brassy yellow color and a metallic luster. Pyrite is frequently found in large, well-formed crystals and is composed of iron(II) and S^{2-}_2 ions in a distorted rock salt structure. Pyrite is unreactive unless heated. It affords Fe_2O_3 and SO_2 in air or FeS and sulfur in a vacuum. Roasting of pyrites has been used in the past to produce SO_2 for sulfuric acid production and iron(III) oxide for use as an iron ore (see SULFURIC ACID AND SULFUR TRIOXIDE). Marcasite is less stable than pyrite and therefore is more reactive. Fe_2S_3 [12063-27-3] is an unstable black precipitate produced when aqueous iron(III) solutions are treated with S^{2-} . It is rapidly decomposed in moist air to Fe_2O_3 and sulfur. A purer material can be obtained by reaction of anhydrous FeCl_3 with bistrimethylsilylsulfide. Iron(III) sulfide occurs in nature in the form of the double sulfide minerals chalcopyrite [1308-56-1], CuFeS_2 , and bornite [1308-82-3], Cu_3FeS_4 , which can be represented as $\text{Cu}_2\text{S} \cdot \text{Fe}_2\text{S}_3$ and $3\text{Cu}_2\text{S} \cdot \text{Fe}_2\text{S}_3$, respectively. Iron(III) sulfide finds applications in cathodes in secondary Li batteries (qv), coal liquefaction, and desulfurization (see COAL CONVERSION PROCESSES).

Chelate Compounds. A chelate is a multidentate ligand which binds to a metal atom at more than one coordination site resulting in a complex having a closed-ring structure (see COORDINATION COMPOUNDS). Chelate complexes are more stable, ie, have greater formation constants, than analogous complexes of unidentate ligands, wherein no rings are formed. The enhanced stability is called the chelate effect and is thought to result predominantly from favorable entropic effects. In general, chelates that contain five-membered rings are more stable than chelates that contain six-membered rings.

Diketones. The protons on the carbon between the two carbonyl groups of 2,4-pentanedione, also called acetylacetone (acacH), and other β -diketone compounds are relatively acidic because the negative charge of the conjugate base anion is delocalized through resonance onto the two carbonyl oxygen atoms. The anion can coordinate to a metal ion by means of the two negatively charged oxygen atoms. This results in formation of a six-membered chelate ring.

In the presence of piperidine, iron(II) sulfate or chloride reacts with 2,4-pentanedione in degassed water under nitrogen to form hydrated bis(2,4-pentanedionato)iron(II). Drying under high vacuum affords the anhydrous compound bis(2,4-pentanedionato)iron(II) [14024-17-0], $\text{Fe}(\text{C}_5\text{H}_7\text{O}_2)_2$ or $\text{Fe}(\text{acac})_2$ (2). Although the stoichiometry suggests the compound is simple, it is coordinatively unsaturated and has an unusual tetrameric structure which consists of two asymmetric Fe_2 linked by long Fe–C bonds. The acac-ligand which contains the methylene carbon bound to iron in the second asymmetric unit also has an oxygen atom that bridges the two iron atoms of the first asymmetric unit. $\text{Fe}(\text{acac})_2$ reacts with numerous bases to form six-coordinate adducts. All of the iron(II) acac complexes are air sensitive. $\text{Fe}(\text{acac})_2$ is used as a catalyst in several types of reactions.



Tris(2,4-pentanedionato)iron(III) [14024-18-1], $\text{Fe}(\text{C}_5\text{H}_7\text{O}_2)_3$ or $\text{Fe}(\text{acac})_3$, forms ruby red rhombic crystals that melt at 184°C. This high spin complex is obtained by reaction of iron(III) hydroxide and excess ligand. It is only slightly soluble in water, but is soluble in alcohol, acetone, chloroform, or benzene. The structure has a near-octahedral arrangement of the six oxygen atoms. Related complexes can be formed with other β -diketones by either direct synthesis or exchange of the diketone into $\text{Fe}(\text{acac})_3$. The complex is used as a catalyst in oxidation and polymerization reactions.

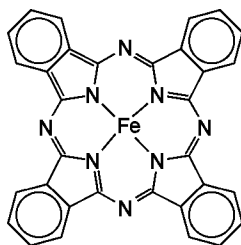
Ethylenediaminetetraacetic Acid. Ethylenediaminetetraacetic acid (EDTAH_4) has six potential donor groups: two nitrogen atoms and four carboxylate groups. If EDTA 4⁻ acts as a hexadentate ligand to a metal, the resulting complex contains five five-membered chelate rings and has a charge that is four less than that of the metal ion.

Iron(II) ethylenediaminetetraacetic acid [15651-72-6], Fe(EDTA) or *N,N'*-1,2-ethanediybis[N-(carboxymethyl)glycinato]ferrate(2⁻), is a colorless, air-sensitive anion. It is a good reducing agent, having $E^\circ = 0.117\text{ V}$, and has been used as a probe of outer sphere electron-transfer mechanisms. It can be prepared by addition of an equivalent amount of the disodium salt, Na₂H₂EDTA, to a solution of iron(II) in hydrochloric acid. Diammonium [56174-59-5] and disodium [14729-89-6] salts of Fe(EDTA)²⁻ are known.

Iron(III) ethylenediaminetetraacetic acid [15275-07-7], Fe(EDTA)³⁻ or *N,N'*-1,2-ethanediybis[N-(carboxymethyl)glycinato]ferrate(1⁻), is a pale yellow, high spin complex in which EDTA serves as a hexadentate ligand. A coordinated water molecule is also present, making the iron atom seven-coordinate with a pentagonal bipyramidal structure. The stability constant for the formation of the complex is 10²⁵. At low pH, a six-coordinate complex is obtained as a result of protonation and decoordination of one carboxylate of the EDTA⁴⁻ ligand. At high pH, μ -oxo species form and eventually precipitate. The ammonium [21265-50-9] and sodium [15708-41-5] salts of Fe(EDTA)³⁻ have been prepared and are used as oxidizing agents, especially in photographic bleaching and fixing preparations. The complexes also find use as oxidation catalysts and as a therapeutic source of iron.

Macrocycles. The complexes of cyclic, *n*-dentate ligands exhibit enhanced stability over the complexes of open-chain, *n*-dentate ligands with the same donor set. This phenomenon is termed the macrocyclic effect and results at least in part from favorable entropic factors. Iron forms complexes with a wide variety of macrocyclic ligands. The majority of these involve macrocycles having four nitrogens as the donor groups, but macrocycles containing other numbers of donors and having oxygen, sulfur, and mixed donor atoms are also known.

Iron(II) phthalocyanine [132-16-1] (3), a green compound, was first prepared by accident during the manufacture of phthalimide. Phthalocyanines are an important group of blue-green pigments that have excellent color intensity, photochemical and thermal stability, and chemical inertness (see PHTHALOCYANINE COMPOUNDS). They find use in dyes, inks, paints, toners, and optical recording media. Iron(II) phthalocyanine is prepared by reductive cyclization of phthalonitrile with finely divided iron in a high boiling solvent such as 1-chloronaphthalene and is purified by sublimation at 450°C under partial vacuum. The iron in the complex has a square planar coordination geometry and an intermediate spin, $S = 1$, ground state. The complex is insoluble in most noncoordinating organic solvents, but dissolves in very strong acids such as sulfuric and chlorosulfonic acids owing to protonation of the basic bridging aza groups. The compound does not dissolve in hot hydrochloric acid, but instead reacts with it to form a material called chloroferic phthalocyanine [14285-56-4] the nature of which is not fully resolved. Iron(II) phthalocyanine forms adducts in coordinating solvents or in the presence of bases, for example phthalocyaninatobis(pyridine)iron [20219-84-5], which can be low spin. Water-soluble iron phthalocyanine complexes are obtained by sulfonating the phenyl residues to obtain tetrasodium phthalocyaninetetrasulfonateferrate [41867-66-7]. Purer materials may be obtained, however, by cyclization of sulfonated phthalic acid or nitrile monomers. Iron(II) phthalocyanine may be reduced by up to four electrons. The complex finds use as a catalyst for a variety of chemical and electrochemical redox reactions.



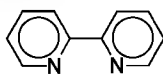
(3)

Oxalates. Iron(II) oxalate dihydrate [6047-25-2], FeC₂O₄·2H₂O, forms by reaction of oxalic acid and aqueous iron(II) solutions. The pale yellow compound is slightly soluble in water, soluble in dilute mineral acids, and decomposes to iron(II) oxide above 190°C. It is used as a photographic developer, to impart greenish brown tints to glass, and as a pigment for plastics, paints, and lacquers. Complex oxalatoferrate(2⁻) salts of general formula M₂[FeC₂(O₄)₂]·*x*H₂O, M monovalent, where *x* depends on M, can be precipitated from solutions which contain alkali oxalates.

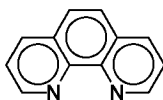
Tris(ethanedioato)ferrate(4⁻) [30948-48-2], [Fe(C₂O₄)₃]⁴⁻, is known but has been studied primarily in solution as an electron-transfer agent.

Tris(ethanedioato)ferrate(3⁻) [15321-61-6], [Fe(C₂O₄)₃]³⁻, can be prepared by addition of an excess of oxalate to a solution of almost any soluble ferric salt. The anion forms isolable salts with many cations including iron(III). Iron(III) oxalate hexahydrate [19469-07-9], Fe₂(C₂O₄)₃·6H₂O, is a yellow powder which is very soluble in water and acid. The green crystalline tripotassium tris(oxalato)ferrate trihydrate [5936-11-8] and triammonium tris(oxalato)ferrate trihydrate [14221-47-7] are among the other known salts. The potassium salt contains high spin iron ions in discrete trisoxalato units. The FeO₆ coordination sphere has some trigonal distortion from octahedral symmetry. The trihydrate yields the anhydrous tripotassium tris(oxalato)ferrate [14883-34-2] at ca 120°C and decomposes at 230°C. Although tris(oxalato)ferrate(3⁻) is stable toward dissociation of oxalate, its solutions and salts are photosensitive. The oxalate ligand is oxidized to CO₂ and iron(III) is concurrently reduced to iron(II). This reaction provides the basis for the first step in the blueprint process. The complex anion is used in other photochemical processes as well as in a variety of redox processes.

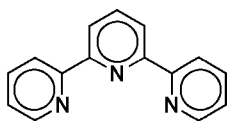
Polypyridyl Ligands. Three important polypyridyl ligands are 2,2'-bipyridine [366-18-7] (bipy) (4), 1,10-phenanthroline [66-71-7] (phen) (5), and 2,2':6',2''-terpyridine [1148-79-4] (terpy) (6). The good σ -donor and π -acceptor properties of these α,α' -diimine chelate ligands make them high field donors of comparable strength to the cyanide ion. As a consequence, the tris chelate complexes of the first two of these ligands and the bis chelate complex of terpy with both iron(II) and iron(III) are low spin. Iron(II) has a very high affinity for these ligands and their substituted derivatives. The complexes are substitution inert and have overall formation constants as high as 10²³. The complexes form readily even in dilute solutions. This and the intense red color of the tris iron(II) or ferroin complexes, where the molar extinction coefficient ϵ_M is ca 10⁴ (M·cm)⁻¹, are responsible for the extensive use of these ligands for the spectrophotometric determination of iron. The large, cationic iron(II) complexes are also useful in the specific precipitation of anions in gravimetric procedures. The phenanthroline complexes find use as reversible, high potential (~1.1 V) redox indicators. The iron(III) complexes are blue and nearly colorless ($\epsilon \sim 10^2$) when compared to the iron(II) complexes. In keeping with the preference of iron(III) for hard donors, the α,α' -diimine chelate complexes of iron(III) have much smaller formation constants. These are unstable with respect to reduction and ligand dissociation. Complexes of the type [Fe(L-L)₂X₂] and [Fe(L-L)X₄]²⁻, where L-L = bipy or phen, are also known.



(4)



(5)



(6)

The tris(2,2'-bipyridine)iron(2+) ion [15025-74-8], [Fe(bipy)₃]²⁺, has an absorption maximum at 522 nm and an absorptivity of 8650 (Mcm)⁻¹. The overall formation constant of the low spin complex is 10¹⁷. For most common counterions the ion has good solubility in water, but may be extracted into organic solvents. The complex dissociates in strong acids and highly alkaline solution. Complexes having different solubility properties can be prepared from bipy ligands which have one or more substituent groups at different ring positions. The symmetry of the FeN unit is D₃ and the five-membered chelate ring is coplanar with the rest of the bipy ligand. Resolution of the enantiomers has been achieved and racemization is slow. The reduction potential of the complex is 1.02 V. Many stable salts have been isolated including the dibromide [15388-40-6], the dichloride [14751-83-8], and the diperchlorate [15388-48-4].

The pale blue tris(2,2'-bipyridine)iron(3+) ion [18661-69-3], [Fe(bipy)₃]³⁺, can be obtained by oxidation of [Fe(bipy)₃]²⁺. It cannot be prepared directly from iron(III) salts. Addition of 2,2'-bipyridine to aqueous iron(III) chloride solutions precipitates the doubly hydroxy-bridged species [(bipy)₂Fe(μ-OH)₂Fe(bipy)₂]Cl₄ [74930-87-3]. [Fe(bipy)₃]³⁺ has an absorption maximum at 610 nm, an absorptivity of 330 (Mcm)⁻¹, and a formation constant of 10¹². In mildly acidic to alkaline aqueous solutions the ion is reduced to the iron(II) complex. [Fe(bipy)₃]³⁺ is frequently used in studies of electron-transfer mechanisms. The triperchlorate salt [15388-50-8] is isolated most commonly.

The orange-red tris(1,10-phenanthroline)iron(2+) ion [14708-99-7], [Fe(phen)₃]²⁺, has an absorption maximum at 510 nm, an absorptivity of 1.10 × 10⁴ (Mcm)⁻¹, and a formation constant of 10²¹. The reduction potential is 1.06 V. The complex is stable in the pH range of 2 to 9, but also persists at higher pH if a reducing agent is present. It has good solubility in water, but may be extracted into organic solvents. The solubility of analogous complex ions in organic solvents increases with the number and size of organic ring substituents. An example is the tris(4,7-diphenyl-1,10-phenanthroline)iron(2+) ion [21412-03-3], which is sometimes used to determine iron concentrations because its higher absorptivity (2.24 × 10⁴ (Mcm)⁻¹ at 533 nm) leads to greater sensitivity. [Fe(phen)₃]²⁺ serves as an electron-transfer mediator, is an excellent indicator in redox titrations, and a constituent of many oscillating reaction systems. Well-known salts of [Fe(phen)₃]²⁺ include the dichloride [14978-15-5], the diiodide [15553-89-6], and the diperchlorate [14586-54-0].

In analogy to the situation for bipyridine, the blue tris(1,10-phenanthroline)iron(3+) ion [13479-49-7], [Fe(phen)₃]³⁺, must be obtained by oxidation of the corresponding iron(II) ion. [Fe(phen)₃]³⁺ has an absorption maximum at 590 nm, an absorptivity of 600 (Mcm)⁻¹, and a formation constant of 10¹⁴. In solutions of pH > 4, this species is reduced to the iron(II) complex. The reduction is instantaneous in alkaline solution. At pH < 2, protons compete with iron(III) for the phenanthroline nitrogens and coordination is incomplete. [Fe(phen)₃]³⁺ is used most often in solution as an oxidant, but the trichloride [40273-22-1] and the triperchlorate monohydrate [20774-81-6] salts have been prepared.

The intensely purple bis(2,2':6',2''-terpyridine)iron(2+) ion [17455-70-8], [Fe(terpy)₂]²⁺, has an absorption maximum at 552 nm, an absorptivity of 1.15 × 10⁴ (Mcm)⁻¹, and a formation constant of 10¹⁹. The reduction potential is 1.13 V. The fit of the ligand is not ideal for an octahedral FeN unit. The N–Fe–N angles are less than 90°. The diperchlorate salt [22079-98-7] precipitates upon addition of perchloric acid to solutions of the ion.

The unstable pale blue-green bis(2,2':6',2''-terpyridine)iron(3+) ion [47779-99-7], [Fe(terpy)₂]³⁺, has been obtained by oxidation of [Fe(terpy)₂]²⁺. It is very unstable with respect to reduction by solvent and ligand dissociation. The perchlorate salt [21536-42-5] has been reported.

Organometallic Compounds

The discussion herein is limited to several simple compounds that are stable, readily available, and used as starting materials for many other organoiron compounds.

Carbonyls. Iron pentacarbonyl [13463-40-6], Fe(CO)₅, is a toxic, yellow-orange, oily liquid which does not react with air at room temperature. It has a musty odor, a relatively high vapor pressure Pa (21 torr) at 20°C, and boils at 103°C. It is readily soluble in benzene, hydrocarbons, and ether, but is insoluble in water. Iron pentacarbonyl is prepared by direct reaction of carbon monoxide with finely divided iron metal at somewhat elevated temperature and pressure. It may be prepared at room temperature and pressure if very highly activated iron metal is used. As a consequence of its method of preparation, iron pentacarbonyl can be a contaminant in coal gas and in carbon monoxide which has been stored at high pressure in steel cylinders. The compound has a trigonal bipyramical structure but is stereochemically nonrigid. Iron pentacarbonyl is slowly converted to iron nonacarbonyl by light. At temperatures above ca 100°C and in the absence of a high CO pressure, iron pentacarbonyl decomposes and produces pure metallic iron. Iron pentacarbonyl has been used as an antiknock agent in gasoline. It burns in air to yield finely divided iron(III) oxide which is suitable for use in pigments and polishing compounds. It also finds use in organic synthesis (11,12).

The reaction chemistry of iron pentacarbonyl is extensive. Soft ligands including phosphines and arsines react with the compound thermally and photochemically to afford CO substitution products of the types Fe(CO)₄L and Fe(CO)₃L₂. Alkynes, alkenes, and dienes form compounds in which the unsaturated hydrocarbon is coordinated to Fe(CO)₄ or Fe(CO)₃ groups. Hard ligands such as hydroxide and amines tend to induce redox disproportionations and afford iron carbonyl anions, iron carbonyl hydrides, and iron carbonyl clusters. Halogens and pseudohalogens oxidatively add to Fe(CO)₅ to produce *cis*-Fe(CO)₄X₂ compounds. Reduction of Fe(CO)₅ using sodium amalgam or sodium benzophenone ketyl in refluxing dioxane affords the pyrophoric compound disodium tetracarbonylferrate [59733-73-2], Na₂[Fe(CO)₄], which is useful in several organic transformations as an acyl anion equivalent (12,13).

Diiron nonacarbonyl [15321-51-4], Fe₂(CO)₉, forms as an impurity in iron pentacarbonyl exposed to light. It is prepared more conveniently by photolysis of solutions of Fe(CO)₅ in cooled acetic acid. The material forms as shiny yellow-gold hexagonal platelets. The compound is sensitive and is best stored at reduced temperature under CO or an inert atmosphere. The crystals darken slowly at low temperature. The decomposition products include Fe₃(CO)₁₂ and finely divided iron, which can make the material pyrophoric. Fe₂(CO)₉ is insoluble in water, ether, and benzene and only slightly soluble in alcohols and acetone. The only solvent in which it has appreciable solubility without reaction is Fe(CO)₅. Diiron nonacarbonyl has a structure of approximately D_{3h} symmetry that consists of two face-sharing octahedra (ie, bridging carbonyl ligands at the three shared vertices) with the iron atoms further linked by an Fe–Fe bond. The reactivity of Fe₂(CO)₉ toward ligands is higher than that of Fe(CO)₅. The reactions of the compound give a variety of products, which frequently can also be obtained from Fe(CO)₅ or Fe₃(CO)₁₂.

Triiron dodecacarbonyl [12088-65-2], Fe₃(CO)₁₂, is prepared by heating Fe₂(CO)₉ to 60°C in a variety of inert organic solvents or by treatment of Fe(CO)₅ with alkali followed by MnO₂, a mild oxidizing agent. The compound forms very dark green monoclinic prismatic crystals, which are soluble in a wide variety of solvents. The pure compound oxidizes slowly in air and is somewhat thermally unstable at room temperature. The structure of triiron dodecacarbonyl consists of a triangle of Fe(CO)₄ units connected by Fe–Fe single bonds, one of which is supported by two CO bridges. Triiron dodecacarbonyl is a useful compound for synthesis of iron carbonyl derivatives because it is more reactive than Fe(CO)₅ and more soluble and stable than Fe₂(CO)₉.

Metalloenes. Bis(cyclopentadienyl)iron or ferrocene [102-54-5], Fe(C₅-H₅)₂, is an air and thermally stable orange solid that sublimes above

100°C and melts at 173°C. It is insoluble in water but dissolves in alcohols, ether, and benzene. Ferrocene can be prepared by numerous methods, including the reaction of cyclopentadienyl anion, $C_5H_5^-$, with anhydrous $FeCl_2$. Its extensive reaction chemistry is notable for the aromaticity of the cyclopentadienyl rings, which readily undergo Friedel-Crafts acylation, alkylation, and metallation. It does not undergo reactions typical of conjugated dienes and resists catalytic hydrogenation. Attempts at direct nitration or halogenation result in oxidation to afford the red-blue dichroic ferricinium ion. The reversibility of the ferrocene–ferricinium couple leads to its use as an internal standard in electrochemistry.

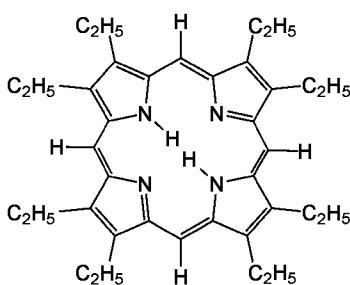
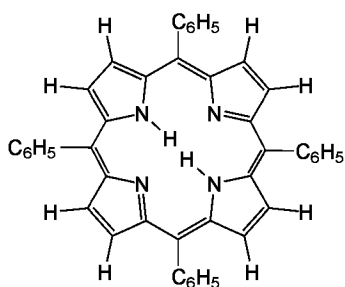
Bis(cyclopentadienyldicarbonyliron) [*12154-95-9*], $[Fe(CO)_2(C_5H_5)]_2$, is a purple-red, air-sensitive solid. It is frequently designated Fp_2 where Fp is an abbreviation for $(C_5H_5)Fe(CO)_2$. The compound is prepared by reacting $Fe(CO)_5$ and dicyclopentadiene at 135°C in an autoclave. Strong reducing agents cleave Fp_2 to Fp in polar aprotic solvents. The anion can be alkylated to afford $(C_5H_5)Fe(CO)_2R$ complexes, which in turn react with hydride abstracting reagents to afford cationic $(C_5H_5)Fe(CO)_2(olefin)$ complexes. Both of these mononuclear compounds have considerable utility in organic syntheses (12,14).

Compounds of Biochemical Relevance

Iron is perhaps the most important of the transition elements that play a role in biochemistry. It is an essential element for all organisms. The functions of iron-containing metalloproteins include electron transfer, dioxygen transport and storage, activation of dioxygen and hydrogen peroxide with concurrent oxidation of substrates, dismutation of superoxide and peroxide, activation and production of dihydrogen, reduction and rearrangement of substrates, and phosphate hydrolysis, among others. Because of the near total insolubility of iron under physiological conditions, iron metalloproteins and chelate compounds function in the solubilization, uptake, transport, and storage of iron.

Iron-containing proteins are classified as either heme proteins or nonheme iron proteins. The former contain iron that is coordinated to a porphyrin ligand.

Iron Porphyrins. Porphyrins (15–17) are aromatic cyclic compounds that consist of four pyrrole units linked at the α -positions by methine carbons. The extended π -systems of these compounds give rise to intense absorption bands in the uv–vis region of the spectrum. The most intense absorption, which is called the Soret band, falls near 400 nm and has $\epsilon_M \sim 10^5$. The π -system is also responsible for the notable ring current effect observed in 1H -nmr spectra, the preference for planar conformations, the prevalence of electrophilic substitution reactions, and the redox chemistry of these compounds. Porphyrins obtained from natural sources have a variety of peripheral substituents and substitution patterns. Two important types of synthetic porphyrins are the meso-tetraaryl porphyrins, such as 5,10,15,20-tetraphenylporphine [*917-23-7*] ($H_2(TPP)$) (7) and β -octaalkylporphyrins, such as 2,3,7,8,12,13,17,18-octaethylporphine [*2683-82-1*] ($H_2(OEP)$) (8). Both types can be prepared by condensation of pyrroles and aldehydes (qv).



Reaction of free-base porphyrin compounds with iron(II) salts in an appropriate solvent results in loss of the two N–H protons and insertion of iron into the tetradentate porphyrin dianion ligand. Five-coordinate iron(III) porphyrin complexes (hemins), which usually have the anion of the iron(II) salt for the fifth or axial ligand, are isolated if the reaction is carried out in the presence of air. Iron(II) porphyrin complexes (hemes) can be isolated if the reaction and workup is conducted under rigorously anaerobic conditions. Typically, however, iron(II) complexes are obtained from iron(III) porphyrin complexes by reduction with dithionite, thiolate, borohydride, chromous ion, or other reducing agents.

Four-coordinate iron(II) porphyrin complexes have the iron atom centered in the plane of the porphyrin, are $S = 1$ intermediate-spin compounds, and can coordinate one or two axial ligands. Five-coordinate iron(II) complexes are square pyramidal where the iron is displaced (up to 50 pm) substantially from the plane. These are high spin, $S = 2$ compounds unless the axial ligand is a strong π -acceptor, in which case the compound has an $S = 0$ spin state. The iron atom moves back toward or into the porphyrin plane in six-coordinate complexes. These remain high spin with weak field axial ligands, but are low spin for stronger field ligands like amines, pyridines, imidazoles, and cyanide. The two stepwise formation constants for coordination of weak field ligands decrease, permitting five-coordinate complexes to be characterized in solution or isolated. Owing to the spin-state change on coordination of a second strong field ligand, the second formation constants for these ligands are usually substantially larger than the first. Thus five-coordinate complexes of strong field ligands are not isolated or observed in solution unless the steric bulk of the ligand precludes formation of a six-coordinate complex.

Iron(II) porphyrins react rapidly with O_2 to afford μ -oxo-bridged complexes $[Fe(II)Por]_2O$ where Por is porphyrin. Antiferromagnetic coupling of the two high spin iron atoms reduces the room temperature magnetic moment to about $1.7 \times 10^{-23} \text{ J/T}$ ($1.8 \mu_B$ Fe). The reaction involves coordination of O_2 to the heme followed by reaction of another equivalent of heme to afford a μ -peroxy-bridged $[Fe(II)Por]_2O_2$ intermediate, homolysis to afford two equivalents of an oxo iron(IV) intermediate, and reaction with yet another equivalent of heme to yield the μ -oxo product. Reversible binding of dioxygen, O_2 , to the heme can occur if steric encumbrance of the O_2 binding site prevents the approach of a second heme. This is the basis of the success of synthetic O_2 binding complexes like picket fence porphyrins. Clearly, one function of the protein surrounding the heme site in the biological oxygen carriers myoglobin and hemoglobin is to isolate the oxygen and prevent its irreversible oxidation. A second function in hemoglobin is to mediate cooperative binding of O_2 by the four heme sites in each molecule. The movement of the iron atom with respect to the porphyrin plane upon O_2 binding is thought to play an important role in cooperativity.

All iron(III) porphyrin complexes are five- or six-coordinate. In five-coordinate complexes the fifth ligand can be one of a variety of anions including halides, carboxylates, alkyls, alkoxides, phenoxides, pseudohalides, and mercaptides among others. Most can be prepared from the conjugate acid of the ligand and the μ -oxo complex or by metathesis with the chloride complex. Typically, five-coordinate complexes are high spin ($S = 5/2$) and have iron displaced roughly 50 pm from the porphyrin plane toward the ligand. The displacement of iron in complexes of weak anionic ligands like ClO_4^- decreases to about 25 pm. As a consequence, the $d_{x^2-y^2}$ orbital is destabilized and the complex adopts an intermediate-spin state. Organometallic

alkyl-iron porphyrin complexes are low spin ($S = 1/2$). Although five-coordinate complexes form adducts with added ligand, the formation constants are relatively small owing to the strong trans-effect of the anionic ligand. Six-coordinate FePorLX complexes can be observed in solution, but generally are not isolable or the sole species in solution. With strong ligands or an excess of ligand, the anionic ligand can be displaced to afford cationic, six-coordinate complexes $[\text{FePorL}_2]\text{X}$, which may be either high or low spin. The iron atom is centered or nearly centered in the porphyrin plane in both cases.

The reduction potential of the iron(III)/iron(II) couple is strongly affected by the number and identity of the axial ligands present. Moreover, spin-state and structural changes, ie, Fe–N and Fe–L bond lengths, and Fe displacement can occur concurrently with electron transfer, which affects the electron-transfer rate. In this light, the variation of axial ligands to the heme group in cytochrome proteins can be interpreted as a mechanism to vary the potential of the protein. The rapidity of the electron transfer in certain cytochromes c such as cytochrome c can be attributed to the low spin nature of both members of the redox couple of the protein and the small changes in structure and bond lengths that therefore occur.

The porphyrin ligand can support oxidation states of iron other than II and III. $[\text{Fe(I)Por}]$ complexes are obtained by electrochemical or chemical reduction of iron(II) or iron(III) porphyrins. The anionic complexes react with alkyl halides to afford alkyl-iron(III) porphyrin complexes. Iron(IV) porphyrins are formally present in the carbene, $\text{RR}'\text{C}=\text{Fe(IV)Por}$; μ -carbido, $\text{PorFe(IV)}-\text{Fe(IV)Por}$; nitrene, $\text{RN}=\text{Fe(IV)Por}$; and μ -nitrido, $\text{PorFe(IV)}-\text{N}=\text{Fe(IV)Por}$ complexes. Oxo iron(IV) porphyrin cation radical complexes, $[\text{O}=\text{Fe(IV)Por}]^+$, are important intermediates in oxygen atom transfer reactions. Compound I of the enzymes catalase and peroxidase have this formulation, as does the active intermediate in the catalytic cycle of cytochrome P_{450} . Similar intermediates are invoked in the extensively investigated hydroxylations and epoxidations of hydrocarbon substrates catalyzed by iron porphyrins in the presence of such oxidizing agents as iodosylbenzene, NaOCl, peroxides, and air.

Iron Sulfur Compounds. Many molecular compounds (18–20) are known in which iron is tetrahedrally coordinated by a combination of thiolate and sulfide donors. Of the 10 or more structurally characterized classes of Fe–S compounds, the four shown in Figure 1 are known to occur in proteins. The mononuclear iron site REPLACE occurs in the one-iron bacterial electron-transfer protein rubredoxin. The $[\text{2Fe}–\text{2S}]$ (10) and $[\text{4Fe}–\text{4S}]$ (12) cubane structures are found in the 2-, 4-, and 8-iron ferredoxins, which are also electron-transfer proteins. The $[\text{3Fe}–\text{4S}]$ voided cubane structure (11) has been found in some ferredoxins and in the inactive form of aconitase, the enzyme which catalyzes the stereospecific hydration–rehydration of citrate to isocitrate in the Krebs cycle. In addition, enzymes are known that contain either other types of iron sulfur clusters or iron sulfur clusters that include other metals. Examples include nitrogenase, which reduces N_2 to NH_3 at a MoFe_7S_8 homocitrate cluster; carbon monoxide dehydrogenase, which assembles acetyl-coenzyme A (acetyl-CoA) at a FeNiS site; and hydrogenases, which catalyze the reversible reduction of protons to hydrogen gas.

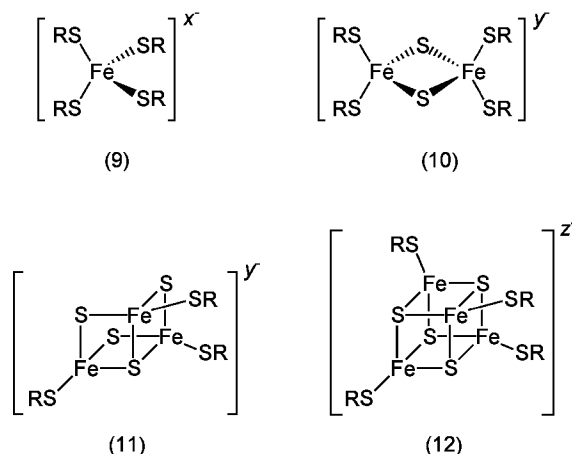


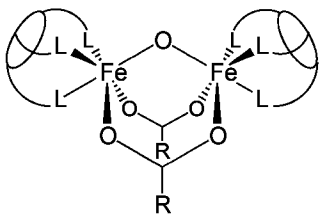
Fig. 1. Four important classes of iron compounds where x , y , and z represent 1 or 2, 2 or 3, and 1, 2, or 3, respectively. Structure (10) is a $2\text{Fe}–\text{2S}$ center; (11), a $3\text{Fe}–\text{4S}$; and (12) a $4\text{Fe}–\text{4S}$.

Low molecular weight complexes that are synthetic analogues of the protein sites have been prepared and extensively investigated in the cases of structures (9), (10), and (12). The compounds, which are typically isolated as tetraalkylammonium salts, assemble spontaneously from a reaction system that includes an iron salt (usually FeCl_3), thiolate, a source of labile S^2 (elemental sulfur or HS^-), and a counterion. Individual compounds can be prepared selectively by variation of the ratios of reactants. Both the iron(II) and iron(III) states of the mononuclear cluster are high spin. Strong intramolecular antiferromagnetic coupling occurs in the cluster compounds, which contain bridging sulfides. Several specific oxidation states of the clusters are mixed valence compounds. $[\text{Fe}_2\text{S}_2(\text{SR})_4]^-$ has a localized iron(II) and a localized iron(III). In contrast, localized iron(II) and iron(III) sites are not observed for the $[\text{4Fe}–\text{4S}]$ compounds, all of which have mixed valent oxidation states. The redox activity of these compounds parallels the states but not the potentials observed in the proteins. Substitution reactions of the terminal thiolate ligands are also noteworthy.

Several iron sulfide nitrosyl compounds are known. These have structures that in some cases are formally related to the FeS clusters by replacement of thiolate by NO. The compounds include the anions $[\text{Fe}_2\text{S}_2(\text{NO})_4]^{2-}$ and $[\text{Fe}_4\text{S}_3(\text{NO})_7]^-$ (Roussin's red and black salts, respectively) and the neutral compounds $[\text{Fe}_2\text{S}_2(\text{NO})_4]$ and $[\text{Fe}_4\text{S}_4(\text{NO})_4]$. Roussin's black salt has found use as a NO releasing vasodilator.

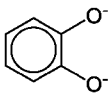
(μ -Oxo)bis(μ -carboxylato) Diiron Complexes. Several nonheme iron proteins of widely varying functions contain a binuclear iron site as a common structural feature. The proteins include hemerythrin, the O_2 -transport protein of marine invertebrates; ribonucleotide reductase, an enzyme which catalyzes the deoxygenation of ribonucleoside diphosphates to deoxyribonucleosides; methane monooxygenase, an enzyme which catalyzes the oxidation of methane to methanol; and purple acid phosphatases, which catalyze the dephosphorylation of phosphoproteins and nucleotides. The site contains two antiferromagnetically coupled iron atoms that are coupled by a bridging oxo or hydroxo group and two bridging carboxylate groups and is recognizable as a portion of the basic ferric acetate structure. The enzymes differ in the nature of the terminal ligands to each iron.

The thermodynamic stability of the binuclear site has been demonstrated by the spontaneous assembly of $[\text{Fe}_2\text{O}(\text{O}_2\text{CR})_2\text{L}_2]$ (13) from ferric salts in the presence of water, an alkyl carboxylate salt, and a tridentate nitrogen donor ligand L that can cap an octahedral face on iron (8). Suitable ligands include tris(pyrazolyl)borates and 1,4,7-triazacyclononanes. Structure (13) is in essence a portion of the basic ferric acetate structure. The complexes are excellent physical and structural models of the diiron sites and model some aspects of reactivity including redox activity and interconversion of the oxo and hydroxo bridge.

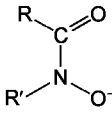


(13)

Siderophores. Iron is not readily available at physiological pH because it is present as the insoluble hydrated iron(III) oxide, which has $K_{sp} \sim 10^{-39}$. Bacteria synthesize chelating agents to facilitate the solubilization of iron from the environment, transport into the organism, and release of iron. Most contain negatively charged oxygen-donor groups which preferentially complex iron(III) and afford octahedral, high spin complexes called siderophores (21,22). The two principal classes of donor groups employed are catecholates (14) and hydroxamates (15).



(14)



(15)

Enterobactin (ent), the cyclic triester of 2,3-dihydroxy-*N*-benzoyl-L-serine, uses three catecholate dianions to coordinate iron. The iron(III)-enterobactin complex [62280-34-6] has extraordinary thermodynamic stability. For $\text{Fe}^{3+} + \text{ent}^{6-}$, the estimated formal stability constant is 10^{40} and the reduction potential is approximately -750 mV at pH 7 (23). Several catecholate-containing synthetic analogues of enterobactin have been investigated and found to have lesser, but still impressively large, formation constants.

Two types of hydroxamate-containing siderophores are known. Ferrichromes are cyclic polypeptides that have three appended hydroxamic acid side chains. Ferrioxamines, which may be linear or cyclic compounds, contain three hydroxamic acids as an integral part of the main backbone rather than as appended side chains. Variants of these compounds, which differ in their substituent groups, can be isolated from different bacteria. The formation constants of these iron(III) complexes typically approach 10^{30} . The free ligand of the complex is designated by adding the prefix deferri- to its name. One of these ligands, deferriferrioxamine B, is marketed as Desferal and is used to treat iron overload from accidental poisoning, chronic transfusions, or Cooley's anemia.

Although high spin iron(III) complexes are usually kinetically labile, siderophores are inert and can be resolved into individual optical isomers. An important issue, then, is how the microorganism releases iron from the siderophore. The redox potential of iron-enterobactin is probably too negative for reduction to iron(II), which would be more labile, to occur under physiological conditions. Suggestions include cleavage of the siderophore backbone and protonation of the chelate groups.

Economic Aspects

Prices of representative iron compounds in 1993 U.S. dollars are listed in Table 1. Suppliers include Aldrich, Alfa Aesar, Cerac, Fisher, and Pressure Chemical, among others.

Table 1. U.S. Prices for Iron Compounds, 1993

Compounds	Formula	CAS Registry Number	Price, ^a \$ /kg
iron(II) acetate	Fe(C ₂ H ₃ O ₂) ₂	[3094-87-9]	1040
iron(III) acetate (basic)	^b	[10450-55-2]	500
iron(III) ammonium citrate	^b	[1185-57-5]	60
iron(II) bromide	FeBr ₂	[7789-46-0]	3000
iron(II) bromide hydrate	FeBr ₂ · xH ₂ O	[13463-12-2]	950
iron(III) bromide	FeBr ₃	[10031-26-2]	1300
iron pentacarbonyl	Fe(CO) ₅	[13463-40-6]	60
diiron nonacarbonyl	Fe ₂ (CO) ₉	[15321-51-4]	1500
triiron dodecacarbonyl	Fe ₃ (CO) ₁₂	[12088-65-2]	1800
iron(II) chloride	FeCl ₂	[7758-94-3]	250
iron(II) chloride tetrahydrate	FeCl ₂ · 4H ₂ O	[13478-10-9]	60
iron(III) chloride	FeCl ₃	[7705-08-0]	12
iron(III) chloride hexahydrate	FeCl ₃ · 6H ₂ O	[10025-77-1]	54
potassium ferrocyanide trihydrate	K ₄ [Fe(CN) ₆] · 3H ₂ O	[14459-95-1]	60
potassium ferrocyanide	K ₄ [Fe(CN) ₆]	[13943-58-3]	33.50
potassium ferricyanide	K ₃ [Fe(CN) ₆]	[13746-66-2]	31.50
sodium pentacyanonitrosylferrate	Na ₂ [Fe(CN) ₅ NO] · 2H ₂ O	[13755-38-9]	176
Prussian Blue (ferric ferrocyanide)	Fe ₄ [Fe(CN) ₆] ₃ ^b	[25869-98-1]	400
iron(II) fluoride	FeF ₂	[7798-28-8]	1400
iron(III) fluoride	FeF ₃	[7783-50-8]	1100
iron(III) fluoride trihydrate	FeF ₃ · 3H ₂ O	[15469-38-2]	640
iron(II) iodide	FeI ₂	[7783-86-0]	2000
iron(II) iodide tetrahydrate	FeI ₂ · 4H ₂ O	[13492-45-0]	750

iron(III) nitrate nonahydrate	Fe(NO ₃) ₃ ·9H ₂ O	[7782-61-8]	18
iron(II) oxalate dihydrate	FeC ₂ O ₄ ·2H ₂ O	[6047-25-2]	28
iron(III) oxalate hexahydrate	Fe ₂ (C ₂ O ₄) ₃ ·6H ₂ O	[19469-07-9]	2400
iron(II) oxide	FeO	[1345-25-1]	1600
iron(III) oxide	Fe ₂ O ₃	[1309-37-1]	10
iron(II,III) oxide	Fe ₃ O ₄	[1317-61-9]	12
iron(III) 2,4-pentanedionate	Fe(C ₅ H ₇ O ₂) ₃	[14024-18-1]	170
iron(II) perchlorate hexahydrate	Fe(ClO ₄) ₂ ·6H ₂ O	[13922-23-8]	36
iron(III) perchlorate hexahydrate	Fe(ClO ₄) ₃ ·6H ₂ O	[13537-24-2]	35
iron(II) phthalocyanine	Fe(C ₃₂ H ₁ N ₈)	[132-16-1]	3200
iron(II) sulfate heptahydrate	FeSO ₄ ·7H ₂ O	[7782-63-0]	42
iron(III) sulfate	Fe ₂ (SO ₄) ₃	[10028-22-5]	50
iron(II) ammonium sulfate	FeSO ₄ ·(NH ₄) ₂ SO ₄ ·6H ₂ O	[7783-85-9]	140
iron(II) sulfide	FeS ^b	[1317-37-9]	860
iron disulfide	FeS ₂	[12068-85-8]	1200
ferrocene	Fe(C ₅ H ₅) ₂	[102-54-5]	120

^a Retail price in U.S. \$ from fine chemical supply houses. Prices may be substantially discounted for larger quantities.
^b Nonstoichiometric compound.

Analytical Methods

Gravimetric. Soluble iron samples may be analyzed by precipitation of iron(III) as the hydrated oxide, followed by heating to 900 °ndash 1000 °degree C to achieve a constant weight of anhydrous Fe₂O₃. The sample must be free of interfering ions, eg, those of aluminum, chromium, titanium, and manganese. The sample is first heated with nitric acid to convert all iron present to iron(III) and then is treated with an excess of ammonia to precipitate hydrated iron(III) oxide as a gelatinous mass. The precipitate is collected on ashless filter paper and is washed with hot 1% ammonium nitrate solution. Paper and precipitate are transferred to a porcelain crucible and ignited.

Volumetric. Analysis of iron solutions requires reduction of all of the iron present to iron(II). This is often accomplished by use of a Jones reductor. The solution of iron(II) may then be titrated with a standardized solution of potassium permanganate, potassium dichromate, cerium(IV) sulfate, or cerium(IV) perchlorate. Titration with permanganate can be complicated by its tendency to oxidize chloride ion, by the temporal instability of permanganate solutions, and by the possibility of uncertain stoichiometry in the reaction. These problems are eliminated by titration with cerium(IV), but it is considerably more expensive, must be used in acidic solution, and requires an indicator such as 1,10-phenanthroline. An indicator, usually diphenylaminesulfonic acid, is also required for titration with potassium dichromate. Dichromate is not as strong an oxidant as either permanganate or cerium(IV) and sometimes may react sluggishly. An advantage of potassium dichromate standard solutions is that they may be prepared directly by weight from the primary standard salt.

Colorimetric. A sensitive method for the determination of small concentrations of dissolved iron is the spectrophotometric determination of the orange-red tris(1,10-phenanthroline)iron(II) complex. Other substituted phenanthrolines can be even more sensitive. Only the iron(II) complexes of these ligands are highly colored. The sample is first treated with an excess of reducing agent. The complexes are stable from pH 2 °ndash 9 and analysis preferably is done at about pH 3.5.

Small concentrations of iron can also be determined by flame atomic absorption and inductively coupled plasma emission spectroscopies (see SPECTROSCOPY).

Health and Safety

Most iron salts and compounds may be safely handled following common safe laboratory practices. Some compounds are irritants. A more serious threat is ingestion of massive quantities of iron salts which results in diarrhea, hemorrhage, liver damage, heart damage, and shock. A lethal dose is 200–250 mg/kg of body weight. The majority of the victims of iron poisoning are children under five years of age.

Two compounds associated with particular industrial risks are iron(III) oxide, Fe₂O₃, and iron pentacarbonyl, Fe(CO)₅. Chronic inhalation of iron(III) oxide leads to siderosis. Adequate ventilation and mechanical filter respirators should be provided to those exposed to the oxide. Iron pentacarbonyl is volatile and highly toxic.

BIBLIOGRAPHY

"Iron Compounds" in *ECT* 1st ed., Vol. 8, pp. 56 °ndash 66, by R. S. Casey, W. A. Scheaffer Pen Co., C. S. Grove, Jr., Syracuse University, and B. J. Lerner, University of Texas; in *ECT* 2nd ed., Vol. 12, pp. 22 °ndash 44, by R. S. Casey, Consultant, and John R. Doyle, State University of Iowa; in *ECT* 3rd ed., Vol. 14, pp. 764 °ndash 788, by J. V. McArdle, University of Maryland.

1. C. M. Flynn, Jr., *Chem. Rev.* **84**, 31 (1984).
2. R. D. Cannon, and R. P. White, *Prog. Inorg. Chem.* **36**, 195 (1988).
3. D. H. R. Barton, M. J. Gastiger, and W. B. Motherwell, *J. Chem. Soc., Chem. Commun.*, 731 (1983).
4. A. G. Sharpe, *The Chemistry of Cyano Complexes of the Transition Metals*, Academic Press, Inc., New York, 1976.
5. H. J. Buser, D. Swartzenbach, W. Petter, and A. Ludi, *Inorg. Chem.* **16**, 2704 (1977); F. Herren, P. Fischer, A. Ludi, and W. Haelg, *Inorg. Chem.* **19**, 956 (1980).
6. R. Colton and J. H. Canterford, *Halides of the First Row Transition Metals*, Wiley-Interscience, New York, 1969.
7. K. B. Yoon and J. K. Kochi, *Inorg. Chem.* **29**, 869 (1990).
8. S. J. Lippard, *Angew. Chem. Int. Ed. Engl.* **27**, 344 (1988).
9. A. F. Wells, *Structural Inorganic Chemistry*, 4th ed., Oxford University Press, Oxford, U.K., 1975, pp. 476 °ndash 515.
10. H. Hibst, *Agnew. Chem. Int. Ed. Engl.* **21**, 270 (1982).
11. H. Alper, in I. Wender and P. Pino, eds., *Organic Syntheses via Metal Carbonyls*, Vol. 2, John Wiley & Sons, Inc., New York, 1977, p. 545.
12. A. J. Pearson, in G. Wilkinson, F. G. A. Stone, and E. W. Abel, eds., *Comprehensive Organometallic Chemistry*, Vol. 8, Pergamon Press, Oxford, U.K., 1982, p. 939.
13. J. P. Collman, *Acc. Chem. Res.* **8**, 342 (1975).
14. D. Astruc, *Tetrahedron* **39**, 4027 (1983).
15. K. M. Smith, ed., *Porphyrins and Metalloporphyrins*, Elsevier Scientific Publishers, Amsterdam, the Netherlands, 1975.
16. D. Dolphin, ed., *The Porphyrins*, Academic Press, Inc., New York, 1979.
17. A. P. B. Lever and H. B. Gray, eds., *Iron Porphyrins*, Vols. 1 and 2, Addison-Wesley, Reading, Mass., 1983; Vol. 3, VCH Publishers, New York, 1989.

18. R. H. Holm, *Acc. Chem. Res.* **10**, 427 (1977).
19. J. M. Berg and R. H. Holm, in T. G. Spiro, ed., *Metal Ions in Biology*, Wiley-Interscience, New York, 1982, Chapt. 1.
20. R. H. Holm, S. Ciurli, and J. A. Weigel, *Prog. Inorg. Chem.* **38**, 1 (1990).
21. J. B. Neilands, *Adv. Chem. Ser.* **162**, 3 (1977).
22. K. N. Raymond and C. J. Carrano, *Acc. Chem. Res.* **12**, 183 (1979); K. N. Raymond, *Coord. Chem. Rev.* **105**, 135 (1990).
23. L. D. Loomis and K. N. Raymond, *Inorg. Chem.* **30**, 906 (1991).

General References

N. N. Greenwood and A. Earnshaw, *Chemistry of the Elements*, Pergamon Press, Oxford, U.K., 1984.
F. A. Cotton and G. Wilkinson, *Advanced Inorganic Chemistry*, 5th ed., John Wiley & Sons, Inc., New York, 1988.
G. Wilkinson, R. D. Gillard, and J. A. McCleverty, eds., *Comprehensive Coordination Chemistry*, Pergamon Press, Oxford, U.K., 1987, Vols. 1,2,4, and 6.
Gmelins Handbuch der Anorganischen Chemie, Springer-Verlag, Berlin, System Number 59.

Alan M. Stolzenberg
West Virginia University

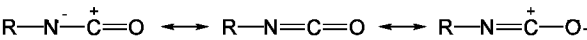


ISOCYANATES, ORGANIC

Isocyanates are derivatives of isocyanic acid, HN=C=O, in which alkyl or aryl groups, as well as a host of other substrates, are directly linked to the NCO moiety via the nitrogen atom. Structurally, isocyanates (imides of carbonic acid) are isomeric to cyanates, ROC≡N (nitriles of carbonic acid), and nitrile oxides, RC≡N→O (derivatives of carboxylic acid).

Isocyanates are liquids or solids which are highly reactive and undergo addition reactions across the C=N double bond of the NCO group. Reactions with alcohols, carboxylic acids, and amines have been widely exploited in developing a variety of commercial products. Cycloaddition reactions involving both the C=N and the C=O double bond of the NCO group have been extensively studied and used for product development (1–9).

The basis for the high reactivity of the isocyanates is the low electron density of the central carbon as indicated by the following resonance structures:



Electron withdrawing or donating substituents alter the electrophilic nature of the isocyanate. Thus, whereas *p*-*N,N*-dimethylaminophenyl isocyanate [16315-59-6] is a rather slow reacting material, sulfonyl or acyl isocyanates are noted to be extremely reactive. The reactivity of isocyanates is also manifested in their tendency to react with themselves to form dimers, trimers, or higher oligomers and polymers. Analytically, isocyanates are readily identifiable through derivatization (urea formation) or via spectroscopy using the strong absorbance between 2300 and 2200 cm^{–1}. Many isocyanates are strong lachrymators (tear-inducing agents). Toxicity data for many of the commercially available isocyanates are discussed herein.

Industrially, isocyanates have become large-volume raw materials for addition polymers, such as polyurethanes, polyureas, and polyisocyanurates. By varying the reactants (isocyanates, polyols, polyamines, and others) for polymer formation, a myriad of products have been developed ranging from flexible and rigid insulation foams to the high modulus automotive exterior parts to high quality coatings and abrasion-resistant elastomers unmatched by any other polymeric material. The most significant mono-, di-, and oligomeric isocyanates, which constitute over 90% of global isocyanate production, are listed in Tables 1–3.

Table 1. Commercially Available Aromatic Isocyanates

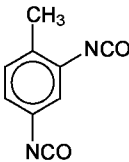
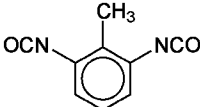
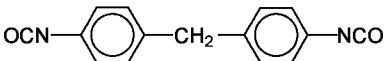
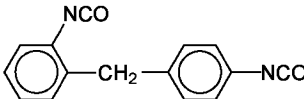
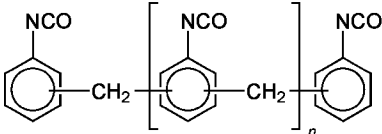
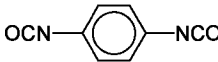
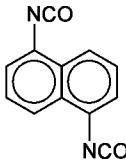
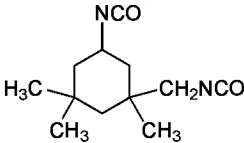
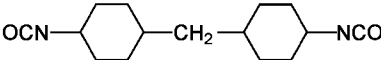
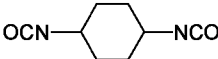
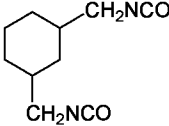
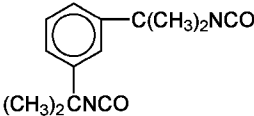
Name	CAS Registry Number	Structure
toluene 2,4-diisocyanate (TDI)	[584-84-9]	
toluene 2,6-diisocyanate (TDI)	[91-08-7]	
4,4'-methylene diphenyl diisocyanate (MDI)	[101-68-8]	
2,4'-methylene diphenyl diisocyanate	[5873-54-1]	
polymeric methylene diphenyl diiso-cyanate (PMDI)	[9016-87-9]	
<i>p</i> -phenylene diisocyanate (PDI)	[935-16-0]	
naphthalene-1,5-diisocyanate (NDI)	[3173-72-6]	

Table 2. Commercially Available Aliphatic Isocyanates

Name	CAS Registry Number	Structure
1,6-hexamethylene diisocyanate (HDI)	[822-06-0]	$\text{OCN}(\text{CH}_2)_6\text{NCO}$
isophorone diisocyanate (IPDI)	[4098-71-9]	
4,4'-dicyclohexylmethane diisocyanate (H_{12} MDI) ^a	[5124-30-1]	
1,4-cyclohexane diisocyanate (CHDI) ^a	[2556-36-7]	
bis(isocyanatomethyl)cyclo-hexane (H XDI,DDI) ^a	[38661-72-2]	
tetramethylxylylene diisocyanate (TMXDI)	[2778-42-9]	

^a Mixture of stereoisomers.

Table 3. Industrially Important Specialty Monoisocyanates

Name	CAS Registry Number
methyl isocyanate (MIC)	[624-83-9]
<i>n</i> -butyl isocyanate (BIC)	[111-36-4]
phenyl isocyanate (PIC)	[103-71-9]
3-chlorophenyl isocyanate	[2909-38-8]
3,4-dichlorophenyl isocyanate	[102-36-3]
<i>p</i> -toluenesulfonyl isocyanate	[4083-64-1]

Synthetic Methods

The first synthetic route for isocyanates was reported in 1848 (10,11). Subsequent efforts by Hofmann, Curtius, and Hentschel pioneered alternative synthetic approaches (12). These efforts highlighted the phosgene–amine approach. Staudinger presented the structural similarities between isocyanates and ketenes and stimulated interest in this class of compounds (13). However, it was not until 1945, when the world was pressed for an alternative to natural rubber, that synthetic routes to isocyanates became an area of great importance. Several excellent review articles covering the synthesis and chemistry of isocyanates have been presented (1–9).

Preparation from Amines. The most common method of preparing isocyanates, even on a commercial scale, involves the reaction of phosgene [75-44-5] and aromatic or aliphatic amine precursors. The initial reaction step, the formation of N-substituted carbamoyl chloride (1), is highly exothermic and is succeeded by hydrogen chloride elimination which takes place at elevated temperatures.



To suppress the formation of side products, primarily ureas and isocyanurates, excess phosgene is employed. In most small-scale batch processes, diluted amine is added gradually to chilled solutions of phosgene followed by venting of excess phosgene. Dehydrohalogenation of the resultant carbamoyl chloride is performed at approximately 80–100°C. Isolation of the desired isocyanate products is achieved by distillation at reduced pressure. The carbamoyl chlorides of certain alkyl isocyanates are stable (or recombine from isocyanate and hydrogen chloride) and distill without decomposing. Because distillation of the crude alkyl isocyanates does not necessarily yield a purified product, alternative synthetic routes are often preferred.

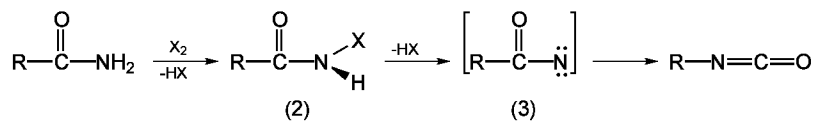
A variation of this method involves the conversion of the amine into the amine hydrochloride prior to treatment with phosgene. This method has the advantage of producing generally cleaner products by retarding the secondary reaction of the free amine with carbamoyl chloride.

The reaction of isocyanate precursors with reactive groups and phosgene also results in isocyanates. For example, *o*-aminobenzoic acid [118-92-3] and phosgene react to form 2-isocyanatobenzoyl chloride [5100-23-2]. Interestingly, isocyanatophenols have been synthesized from aminophenols, under controlled conditions, without formation of the corresponding chloroformates (14,15).

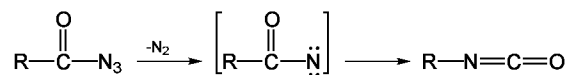
Instead of amines, sulfonamides have also been used as starting materials, producing the highly reactive sulfonyl isocyanates, which have found applications in the manufacture of drugs for diabetics and as drying agents (qv) (4,9,16).

Oligomers of phosgene, such as diphosgene [503-38-8], $(\text{COCl}_2)_2$, have found use in the laboratory preparation of isocyanates. Carbamoyl chlorides, *N,N*-disubstituted ureas, dimethyl- and diphenylcarbonates, and arylsulfonyl isocyanates have also been used to convert amines into urea intermediates, which are subsequently pyrolyzed to yield isocyanates. These methods have found applications for preparation of low boiling point aliphatic isocyanates (2,9,17).

Preparation from Nitrene Intermediates. A convenient, small-scale method for the conversion of carboxylic acid derivatives into isocyanates involves electron sextet rearrangements, such as the ones described by Hofmann and Curtius (12). For example, treatment of benzamide [55-21-0] with halogens leads to an *N*-haloamide (2) which, in the presence of base, forms a nitrene intermediate (3). The nitrene intermediate undergoes rapid rearrangement to yield an isocyanate. Ureas can also be formed in the process if water is present (18,19).

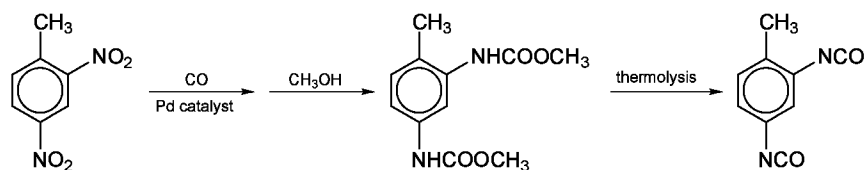


More convenient is the use of aryl azides which are readily converted into isocyanates upon heating in nonreactive solvents via the loss of nitrogen. The latter method is useful for the synthesis of isocyanates with additional substituents which could not be prepared with phosgene (20).

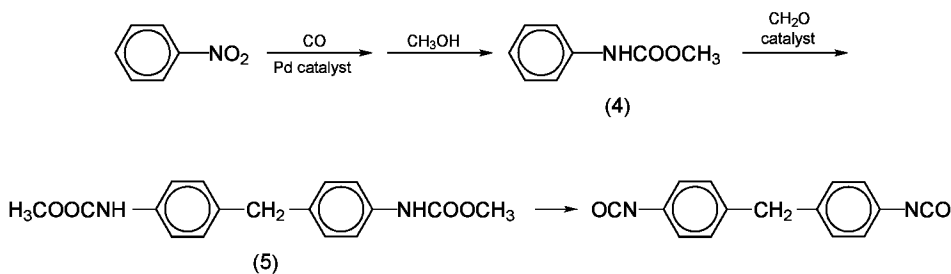


A process for the commercial synthesis of *p*-phenylene diisocyanate using terephthalamide [3010-82-0] as a precursor and involving *N*-halo intermediates has been studied extensively (21). The synthesis of 1,4-diisocyanatocyclohexane from terephthalic acid [100-21-0] also involves a nitrene intermediate (22).

Nonphosgene Preparation. The term nonphosgene route is primarily used in conjunction with the conversion of amines (or the corresponding nitro precursor) to isocyanates via the use of carbonylation agents. These multistep approaches are becoming more attractive to the chemical industry as environmental or toxicological restrictions involving chlorine or phosgene are increasingly enforced. For example, 2,4-dinitrotoluene [121-14-27] undergoes reductive carbonylation to form 2,4-toluene diisocyanate (TDI) in the presence of palladium catalysts (23–27). A variation of this process consists in capturing the isocyanate formed with methanol followed by thermolysis of the biscarbamate (26).



Similarly, nitrobenzene, carbon monoxide, and methanol can react sequentially in the presence of noble metal catalysts, to produce methyl *N*-phenylcarbamate [2603-10-3] (4). The phenylcarbamate is subsequently coupled with formaldehyde [50-00-0] to yield the methylenebis(carbamate) (5) which is pyrolyzed to yield methylene diphenyl diisocyanate (MDI) (23).



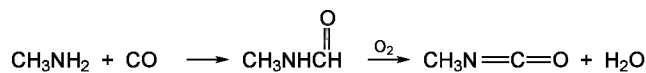
Both dimethyl carbonate [616-38-6] and diphenyl carbonate [102-09-0] have been used, in place of carbon monoxide, as reagents for the conversion of amines into isocyanates via this route (28,29). Alternatively, aniline [62-53-3], toluene diamines (TDA), and methylene dianilines (MDA) have also been used as starting materials in the carbonylations to provide a wide variety of isocyanate monomers.

A simpler nonphosgene process for the manufacture of isocyanates consists of the reaction of amines with carbon dioxide in the presence of an aprotic organic solvent and a nitrogenous base. The corresponding ammonium carbamate is treated with a dehydrating agent. This concept has been applied to the synthesis of aromatic and aliphatic isocyanates. The process relies on the facile formation of amine–carbon dioxide salts using acid halides such as phosphoryl chloride [10025-87-3] and thionyl chloride [7719-09-7] (30).

More recently, a commercial process has been introduced for the manufacture of methyl isocyanate (MIC) which involves the dehydrogenation of *N*-methylformamide [123-39-7] in the presence of palladium, platinum [7440-06-4], or ruthenium [7440-18-8], at temperatures between 50–300°C (31). Aprotic solvents, such as benzene [71-43-2], xylenes, or toluene [108-88-3], may optionally be used. A variation of this synthesis employs stoichiometric amounts of palladium chloride [7647-10-1], PdCl_2 .



Du Pont has reported an alternative catalytic process for the production of MIC starting with methylamine [74-89-5] (32).

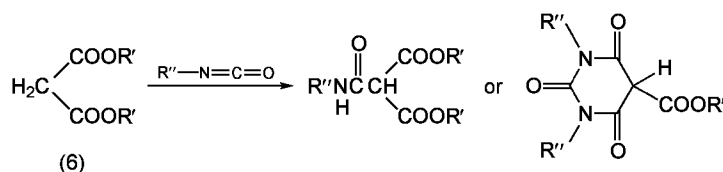


The above processes are only selected examples of a vast number of process options. In the case of carbonylation, the formation of by-products, primarily isocyanate oligomers, allophanates, and carbodimides, is difficult to control and is found to greatly reduce the yield of the desired isocyanate. Thus a number of nonphosgene processes have been extensively evaluated in pilot-plant operations, but none have been scaled up to commercial production of diisocyanates primarily due to process economics with respect to the existing amine–phosgene route. Key factors preventing large-scale commercialization include the overall reaction rates and the problems associated with catalyst recovery and recycle.

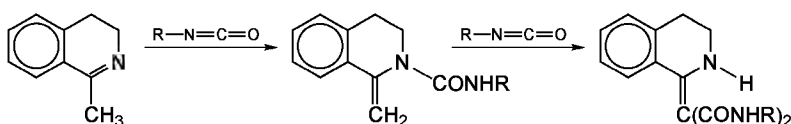
Chemical Properties

Addition Reactions. Isocyanates undergo addition reactions with a wide variety of substrates. Preferred addition occurs across the C=N bond of the NCO moiety. In general, the electron-poor carbon of the isocyanate group serves as the center for attack of isocyanates on electron-rich centers. Electron-withdrawing groups linked to the isocyanate group increase the reactivity of the NCO group. The most commonly used reaction of isocyanates involves their addition to alcohols, amines, and carboxylic acids. Translated to di- or polyfunctional starting materials, these products represent the basis for the diverse group of products generally referred to as polyurethanes.

Insertion Reactions. Isocyanates also may undergo insertion reactions with C—H bonds. Acidic compounds, such as 1,3-dicarbonyl compounds (6), react readily at room temperature to form carboxyamides. At higher temperatures carboxyamides frequently undergo secondary reactions leading to cyclized products (33,34).

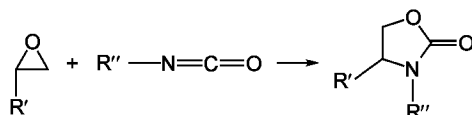


Another common process involves reaction with C=C or C=N species having adjacent CH₂ or CH₃ groups. Initial attack of the isocyanate is on the electron-rich center of the double bond with subsequent migration and insertion of the CONR group into the CH bond. Suitable reagents include *N*-alkylated acetamidines, 1-methyl dihydroisoquinoline, and 2-methyl-2-oxazoline [1120-64-5] (35).

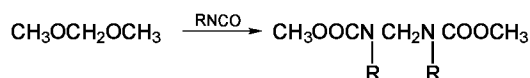


A variety of olefins or aromatic compounds having electron-donating substituents are known to undergo C—H insertion reactions with isocyanates to form amides (36,37). Many of these reactions are known to involve cyclic intermediates.

Isocyanates insert into RO and RN bonds. Cyclic ethers, such as oxiranes, are known to undergo reactions with isocyanates to form 2-oxazolidinones in high yield (38–40).



Similarly, dimethoxymethane or cyclic acetals react to form carbamates in the presence of catalysts.



Tertiary amines have been shown to react with isocyanates in an analogous fashion to form ureas (41–43). Similarly, aziridines (three-membered rings containing nitrogen) are found to react with isocyanates to yield cyclic ureas. Tertiary amines have also been shown to form labile dipolar 1:1 adducts with isocyanates reminiscent of salt formation. In contrast, formaldehyde *N,N*-acetal aminals form insertion products with sulfonyl isocyanates (44,45).

Cycloaddition Reactions. Isocyanates undergo cycloadditions across the carbon–nitrogen double bond with a variety of unsaturated substrates. Addition across the C=O bond is less common. The propensity of isocyanates to undergo cyclization reactions has been widely explored for the synthesis of heterocyclic systems. Substrates with C=O, C=N, C=S, and C=C bonds have been found to yield either 2 + 2, 2 + 2 + 2, or 2 + 4 cycloadducts or a variety of secondary reaction products (2).

Most reactions of this type were found to involve acyclic 1,4-dipolar intermediates which cyclize to four-membered heterocycles or are intercepted by isocyanate or C=X components, such as C=N, C=S, and CR₂, to form a six membered ring. This group of reactions is illustrated in Figure 1. Depending on the nature of the isocyanate and the double-bond system, any of the products shown in Figure 1 can be obtained. Variations in the component ratio or judicious choice of reagents are noted to have pronounced control of product type. Additional reaction details, as well as a description of the multiple transformations involving adjacent substituents, have been summarized (28).

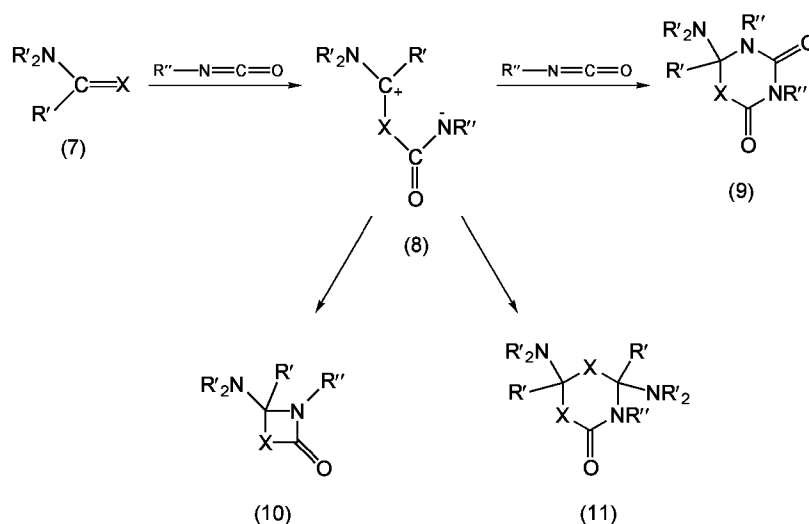
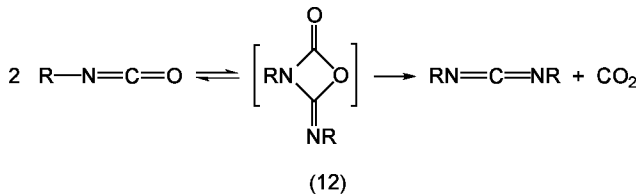
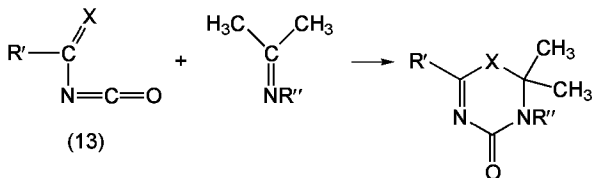


Fig. 1. Cycloaddition reactions of isocyanates: (7) X = NR, S, O, or CR₂; (8) 1,4-dipolar intermediate; (9) 2:1 adduct; (10) 1:1 cycloadduct; and (11) 1:2 adduct.

The dimerization and trimerization of isocyanates are special cases of the cycloaddition reaction in that they involve reagents of the same type. The uncatalyzed carbodiimidization of isocyanates likely involves a labile 2 + 2 cycloadduct (12) which liberates carbon dioxide.



Acyl isocyanates (13,X = O,S) have been shown to react as heterodienes in most cycloadduct formations. Notable examples include autodimerization and the addition to imines (46,47). Unlike aromatic isocyanates, it is not possible to predict the reaction pathway nor the structure of the products which may arise from a given approach or set of reaction conditions.



A large number of Diels-Alder-type reactions, involving both aromatic and sulfonyl isocyanates, have been reported. Heterodienes having high electron density are found to add to isocyanates to form six membered heterocycles as shown in Figure 2 (48–50).

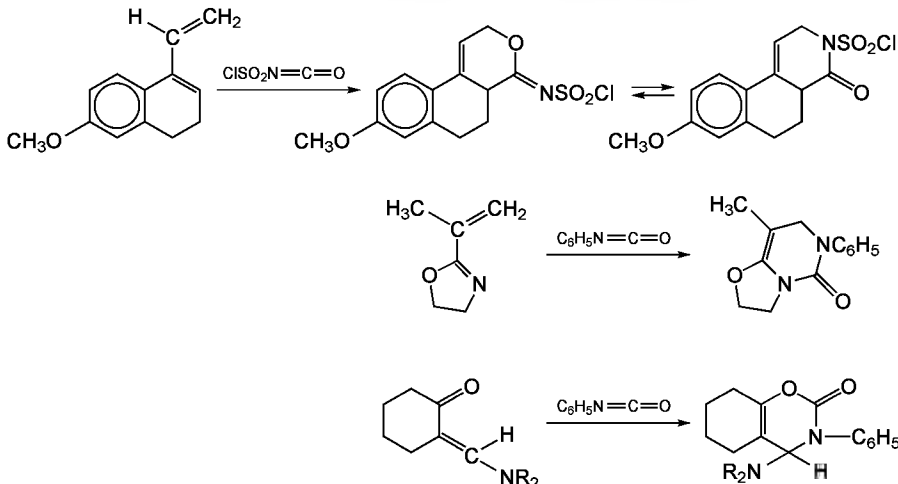
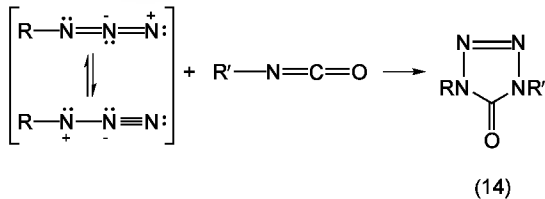
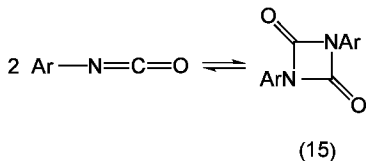


Fig. 2. Diels-Alder-type reactions of aromatic and sulfonyl isocyanates.

A comprehensive review of reactions of isocyanates and 1,3-dipolar compounds has been previously published (51). The example shown illustrates the reaction of azides and isocyanates to yield tetrazoles (14,R = alkyl or aryl, R' = aryl or sulfonyl) (52,53).

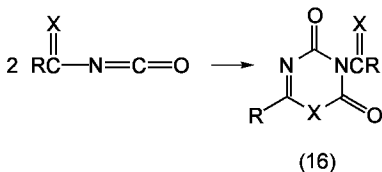


Oligomerization and Polymerization Reactions. One special feature of isocyanates is their propensity to dimerize and trimerize. Aromatic isocyanates, especially, are known to undergo these reactions in the absence of a catalyst. The dimerization product bears a strong dependency on both the reactivity and structure of the starting isocyanate. For example, aryl isocyanates dimerize, in the presence of phosphorus-based catalysts, by a crosswise addition to the C=N bond of the NCO group to yield a symmetrical dimer (15).



Slow dimerization is generally noted to occur in some isocyanates during prolonged storage. The tendency for 4,4'-methylene diphenyl diisocyanate to undergo uncatalyzed dimerization is tied to its crystal structure. The molecules of the 4,4'-MDI align in the solid state, with the NCO groups in close proximity, which leads to slow formation of the dimer at room temperature. The structure of the symmetrical MDI dimer has been verified by x-ray analysis (54). It has been reported that substituted benzyl isocyanates form mixtures of both dimers and trimers in high yield when 1,2-dimethylimidazole [1739-84-0] is used as a catalyst (55).

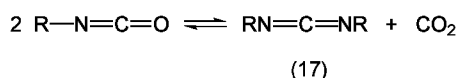
Conversely, acyl isocyanates yield dimers which include the C=X (16) moiety (where X = O, S, NR) in the product ring structure (56).



Reportedly, simple alkyl isocyanates do not dimerize upon standing. They trimerize to isocyanurates under comparable reaction conditions (57). Aliphatic isocyanate dimers can, however, be synthesized via the phosgenation of *N,N'*-disubstituted ureas to yield *N*-(chlorocarbonyl)chloroformamidine intermediates which are subsequently converted by partial hydrolysis and base catalyzed cyclization. This is also the method of choice for the synthesis of 1-alkyl-3-aryl-1,3-diazetidiones (mixed dimers of aromatic and aliphatic isocyanates) (58).

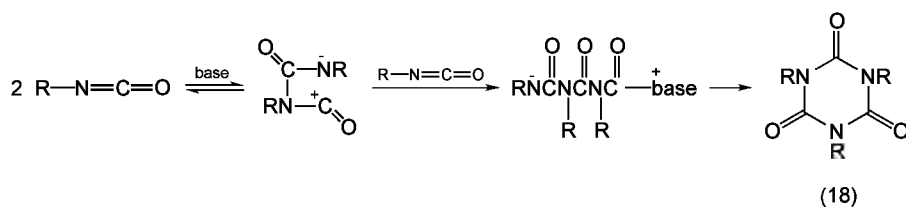
Dimerization is reportedly catalyzed by pyridine [110-86-1] and phosphines. Trialkylphosphines have been shown to catalyze the conversion of dimer into trimer upon prolonged standing (2,57). Pyridines and other basic catalysts are less selective because the required increase in temperature causes trimerization to compete with dimerization. The gradual conversion of dimer to trimer in the catalyzed dimerization reaction can be explained by the assumption of equilibria between dimer and polar catalyst–dimer intermediates. The polar intermediates react with excess isocyanate to yield trimer. Factors, such as charge stabilization in the polar intermediate and its lifetime or steric requirement, are reported to be important. For these reasons, it is not currently feasible to predict the efficiency of dimer formation given a particular catalyst.

Asymmetric aryl isocyanate dimers, in which the C=O group of one molecule reacts with the C=N group of another, have been postulated as labile intermediates in the formation of carbodiimides (17) upon heating isocyanates.



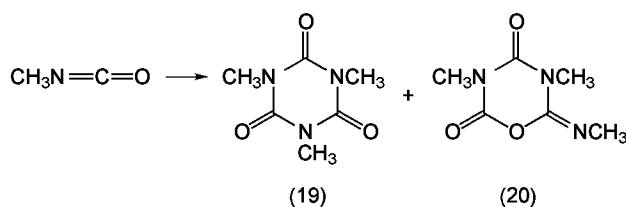
Carbodiimide formation is markedly accelerated when phosphine oxides (R_3PO) or phosphates are used as catalysts. Intermediates with $\text{P}-\text{NR}$ bonds have been postulated as intermediates in these reactions (59,60).

Both alkyl and aryl isocyanates are found to trimerize upon heating or in the presence of catalysts to 1,3,5-trisubstituted hexahydro-*s*-triazinettriones (18) (isocyanurates) (57). Only highly substituted isocyanates, such as *tert*-butyl isocyanate [7188-38-7] and *tert*-octyl isocyanate, fail to trimerize under these conditions.



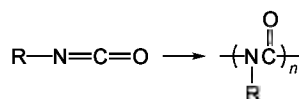
Commercially, polymeric MDI is trimerized during the manufacture of rigid foam to provide improved thermal stability and flammability performance. Numerous catalysts are known to promote the reaction. Tertiary amines and alkali salts of carboxylic acids are among the most effective. The common step in all catalyzed trimerizations is the activation of the C=N double bond of the isocyanate group. The example (18) highlights the alkoxide assisted formation of the cyclic dimer and the importance of the subsequent intermediates. Similar oligomerization steps have been described previously for other catalysts (61).

Interestingly, methyl isocyanate is noted to form unusual trimer products in the presence of trialkylphosphine catalysts. Both the expected triazine (19) and 3,5-dimethyl-2-methylimino-4,6-dioxohexahydro-1,3,5-oxadiazine (20) products are formed (62).



Diisocyanates undergo anionic homopolymerization at subambient temperatures in polar solvents to yield high molecular weight cross-linked isocyanates. This type of polymerization has generally been observed for short-chain, aliphatic diisocyanates which are structurally conducive to an alternating intermolecular and intramolecular propagation mechanism. The thermal homopolymerization of 2,4-toluene diisocyanate [584-84-9] has been reported. The structure of the resultant low molecular weight oligomers has not been established.

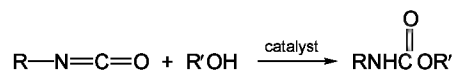
Monoisocyanates undergo anionic homopolymerization at subambient temperatures to yield nylon-1 polymers (polyamides) (63).



Although the crystalline poly(allyl isocyanate) polymers are reported to be stable, many of these polymers depolymerize upon heating to yield monomers and cyclic trimers. The level of temperature sensitivity is a strong function of the length of the side chain. Room temperature depolymerization occurs in polar solvents in the presence of an initiator. Interestingly, the solution properties of poly(alkyl isocyanates) display an unusual degree of chain stiffness which is attributed to their helical configuration (64).

Addition Polymers. The most commonly referenced reaction of isocyanates involves their addition to polyhydroxyl, polyamine, or polycarboxylic acid compounds to yield addition polymers. Due to the wide diversity of raw material characteristics and the broad range of functionality, polyurethane polymers having a wide range of processing and performance characteristics are available.

The reaction of isocyanates with alcohols to form carbamates is catalyzed by amines and a variety of organometallic compounds.

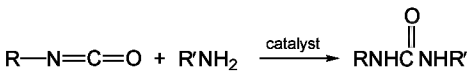


This simple reaction is the bedrock of the polyurethane industry (see URETHANE POLYMERS). Detailed descriptions of the chemistry and process have been published (65–67). Certain carbamates are known to reversibly yield the isocyanate and polyol upon heating. This fact has been commercially used to synthesize a number of blocked isocyanates for elastomer and coating applications.

Similarly, thioalcohols and thiophenols react with isocyanates to form thiocarbamates. Although these reactions are generally found to be much slower than that of the corresponding alcohol, alkoxide catalysts have successfully been used to provide moderate levels of rate enhancement (68).

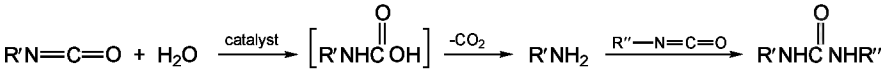


Conversely, the rate of reaction of isocyanates with amines to yield ureas is both rapid and quantitative. Much has been written concerning the reaction



kinetics, solvent effects, and catalysis of this reaction (65–67). The rate of reaction is a strong function of the basicity of the amine. Commercially, this relationship has been used to develop a wide variety of sterically hindered or electronically deactivated aromatic diamine chain extenders for reaction injection molding (RIM) and elastomer applications (see PLASTICS PROCESSING) (69).

Industrially, polyurethane flexible foam manufacturers combine a version of the carbamate-forming reaction and the amine-isocyanate reaction to provide both density reduction and elastic modulus increases. The overall scheme involves the reaction of one mole of water with one mole of isocyanate to produce a carbamic acid intermediate. The carbamic acid intermediate spontaneously loses carbon dioxide to yield a primary amine which reacts with a second mole of isocyanate to yield a substituted urea.

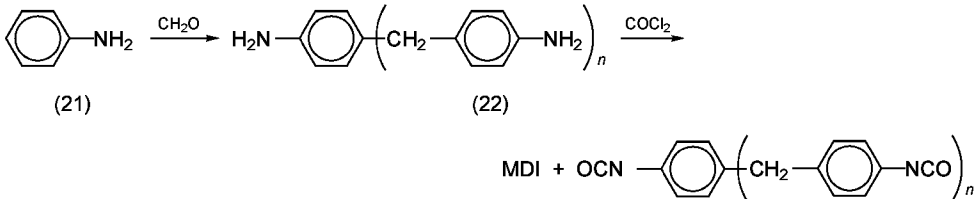


Carboxylic acids react with aryl isocyanates, at elevated temperatures to yield anhydrides. The anhydrides subsequently evolve carbon dioxide to yield amines at elevated temperatures (70–72). The aromatic amines are further converted into amides by reaction with excess anhydride. Ortho diacids, such as phthalic acid [88-99-3], react with aryl isocyanates to yield the corresponding *N*-aryl phthalimides (73). Reactions with carboxylic acids are irreversible and commercially used to prepare polyamides and polyimides, two classes of high performance polymers for high temperature applications where chemical resistance is important. Base catalysis is recommended to reduce the formation of substituted urea by-products (74).

Commercial Manufacturing Processes

Aromatic Isocyanates. A variety of methods are described in the literature for the synthesis of aromatic isocyanates. Only the phosgenation of amines or amine salts is used on a commercial scale (5). Much process refinement has occurred to minimize the formation of disubstituted ureas arising by the reaction of the generated isocyanate with the amine starting material. A listing of the key commercially available isocyanates is presented in Table 1.

For methylene diphenyl diisocyanate (MDI), the initial reaction involves the condensation of aniline [62-53-3] (21) with formaldehyde [50-00-0] to yield a mixture of oligomeric amines (22, where $n = 1, 2, 3, \dots$). For toluene diisocyanate, amine monomers are prepared by the nitration (qv) of toluene [108-88-3] and subsequent hydrogenation (see AMINES BY REDUCTION). These materials are converted to the isocyanate, in the majority of the commercial aromatic isocyanate phosgenation processes, using a two-step approach.



In the first step, a solution of amine is mixed with a solution of phosgene. An excess of phosgene is needed to retard by-product formation. The solvents most commonly used in the phosgenation reaction include toluene, xylene, halobenzenes, and decahydronaphthalene [91-17-8]. The halobenzenes are preferred because of their polarity. In the second step, the resulting carbamoyl chloride-amine hydrochloride slurry reacts with excess phosgene at temperatures in excess of 100°C to yield the isocyanate. The appearance of a clear solution signals the end of the reaction. Distillation of the solvent followed by fractional crystallization, fractional distillation, or sublimation affords pure isocyanate. Typical process flow sheets for the phosgenation of toluene diamine (TDA) and polymeric methylene dianiline (PMDA), respectively, are shown in Figures 3 and 4.

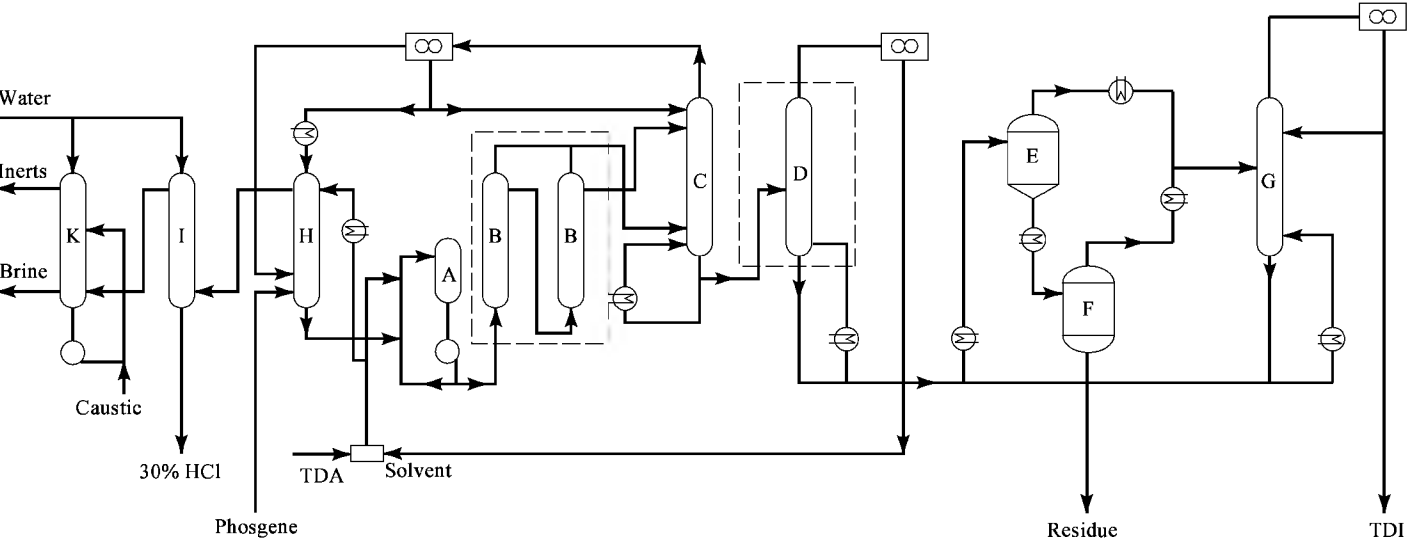


Fig. 3. Schematic of toluene diamine phosgenation process: A, cold phosgenator; B, hot phosgenator; C, wash column; D, solvent distillation; E, preflasher; F, evaporator; G, TDI distillation; H, phosgene removal; I, HCl absorber; and K, phosgene decomposition.

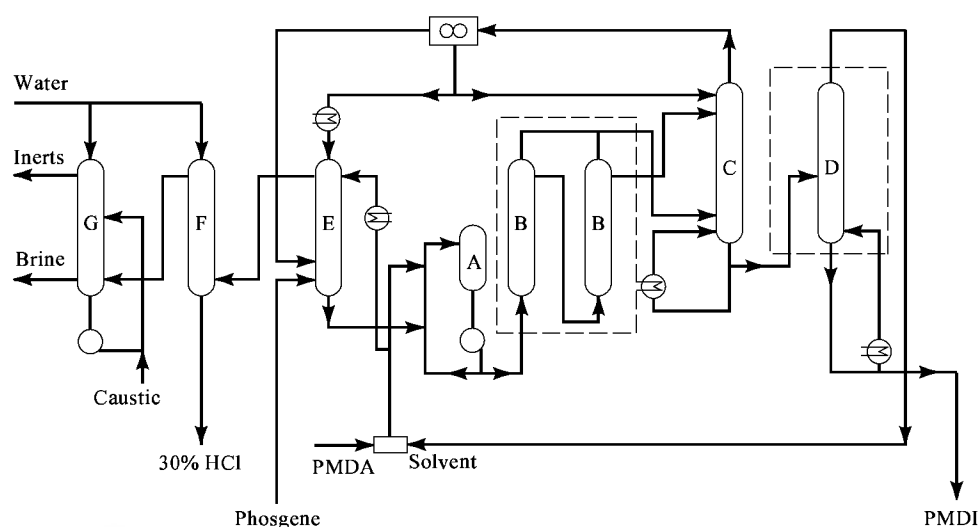
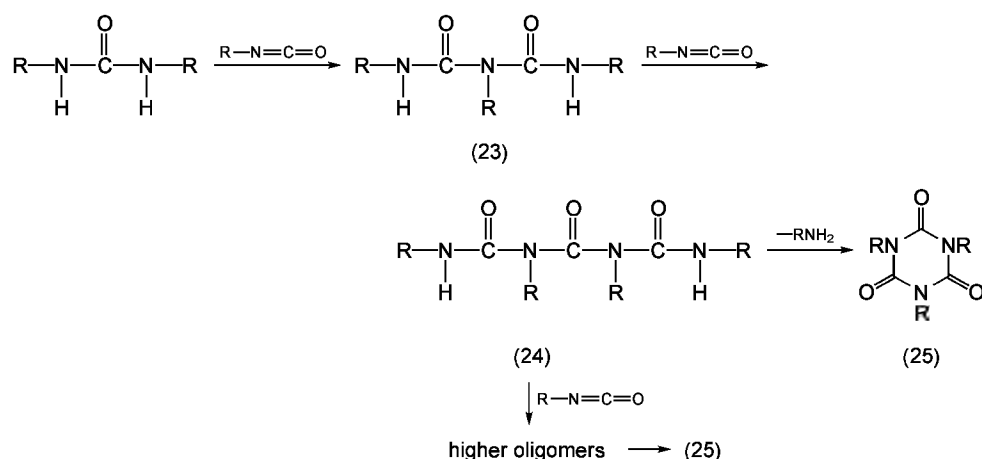


Fig. 4. Schematic of polymeric methylene dianiline phosgenation process: A, cold phosgenator; B, hot phosgenator; C, wash column; D, solvent distillation; E, phosgene removal; F, HCl absorber; and G, phosgene decomposition.

An excess of phosgene is used during the initial reaction of amine and phosgene to retard the formation of substituted ureas. Ureas are undesirable because they serve as a source for secondary product formation which adversely affects isocyanate stability and performance. By-products, such as biurets (23) and triurets (24), are formed via the reaction of the labile hydrogens of the urea with excess isocyanate. Isocyanurates (25, R = phenyl, tolyl) may subsequently be formed from the urea oligomers via ring closure.



These oligomerization steps result in a continuous increase in viscosity of the desired isocyanate and ultimately cause solidification.

The *in situ* generated disubstituted ureas (26) also react with phosgene to yield thermally unstable allophanoyl chlorides (27) and chloroformamides (28) (75). As shown in Figure 5, the allophanoyl chlorides eliminate hydrogen chloride to form the isocyanate. The chloroformamides, however, yield chloroformamidine-*N*-carbonyl chloride (29), which decomposes to yield both carbodiimides (30) and isocyanide dichlorides (31). The carbodiimides simply contribute to yield loss. The isocyanide dichlorides, although present in small amounts, are a contributor to chlorine-containing impurities which detrimentally affect product performance.

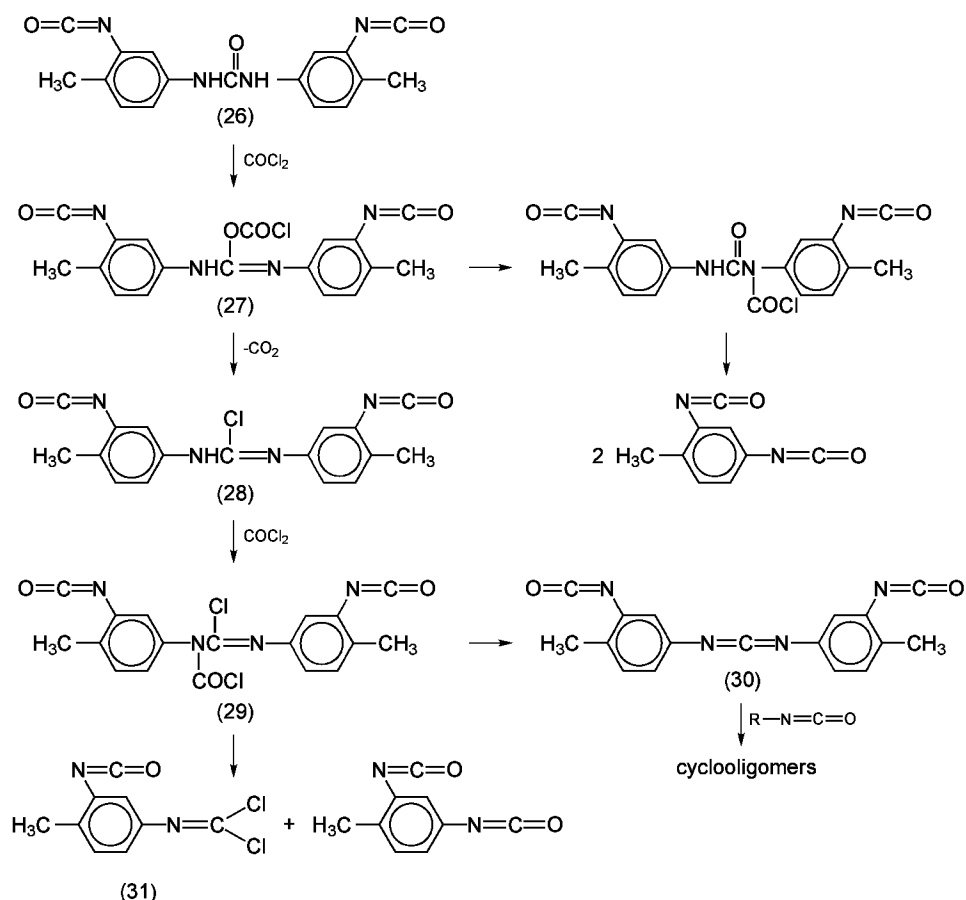
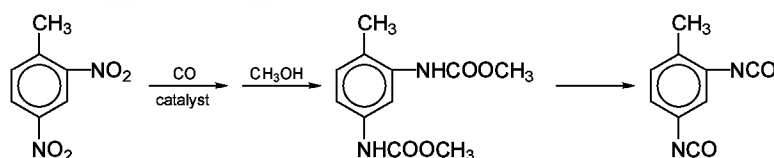


Fig. 5. Reactions of disubstituted ureas.

Alternatively, the aromatic amine can first be treated with hydrogen chloride to form a slurry of amine salts, which is subsequently phosgenated. The slurry is processed using a temperature staged reaction sequence. Excess hydrogen chloride and phosgene are vented to retard the formation of isocyanate recombination products. The isocyanate is purified via solvent separation and fractionation. This method has the disadvantage that gaseous phosgene reacts very slowly with the suspended amine salt; thus, high temperatures and pressures are generally needed.

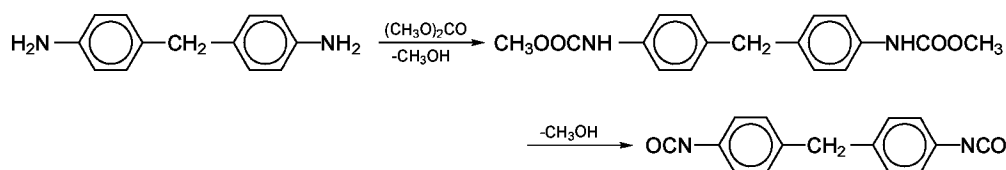
Some of these isocyanates are commercially available in derivatized form, such as biurets and carbodiimides, to provide materials having improved handling or processing characteristics.

Attempts have been made to develop methods for the production of aromatic isocyanates without the use of phosgene. None of these processes is currently in commercial use. Processes based on the reaction of carbon monoxide with aromatic nitro compounds have been examined extensively (23,27,76). The reductive carbonylation of 2,4-dinitrotoluene [121-14-2] to toluene 2,4-dialkylcarbamates is reported to occur in high yield at reaction temperatures of 140–180°C under 6900 kPa (1000 psi) of carbon monoxide. The resultant carbamate product distribution is noted to be a strong function of the alcohol used. Mitsui-Toatsu and Arco have disclosed a two-step reductive carbonylation process based on a cost effective selenium catalyst (22,23). A typical reaction sequence for the production of TDI is as follows.



Two-step approaches based on cocatalysts or alternate catalysts and one-step approaches which circumvent the formation of the bis-carbamate intermediates have also been reported (76–81).

Other approaches have explored the reaction of amines with dimethyl carbonate or its precursors (28). A reaction scheme for the production of polymeric MDI is as follows:



This approach is complicated by the fact that the isocyanate is produced via the thermolytic cleavage of the methyl carbamate. Reactions with the unconverted carbamate cannot be prevented. Much effort has been focused on improving the selectivity of the latter step.

Aliphatic Isocyanates. Conventional aliphatic isocyanates have historically been manufactured using the hydrogen chloride salt slurry approach. Exceptions to this are the longer chain aliphatics which, due to the increased solubility, have reaction rates conducive to the free amine process (82). In the hydrogen chloride salt approach, a fine slurry of salt reacts with phosgene in an agitated autoclave. The reaction must be carried out at temperatures below 150°C to avoid the formation of chlorinated monoisocyanates as anhydrous hydrogen chloride has been found to displace isocyanate groups. Similar to the aromatic isocyanate processes described, the carbamate salt suspension is treated with phosgene using a series of reactors (9). Typically, the amine salt–phosgene reaction is carried out at a temperature of 30°C for 12–24 hours, then finished using a series of 100°C digestion steps. The resultant isocyanate solution is purified by solvent stripping followed by fractional vacuum distillation. A typical process flow sheet for the phosgenation of hexamethylene diamine [124-09-4] is shown in Figure 6.

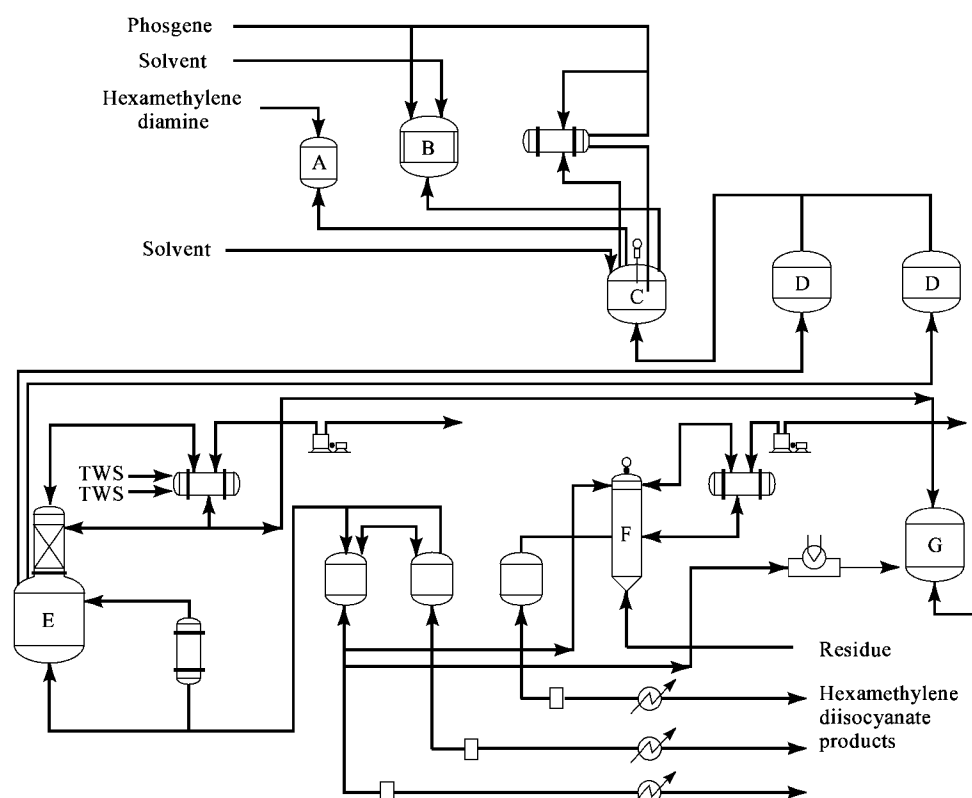
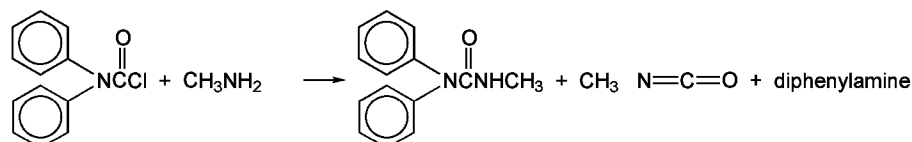


Fig. 6. Schematic of hexamethylene diamine phosgenation process: A, HMDA tanks; B, phosgene solution tanks; C, phosgenation reactor; D, secondary reactors; E, batch still; F, thin-film evaporator; and G, solvent receiver.

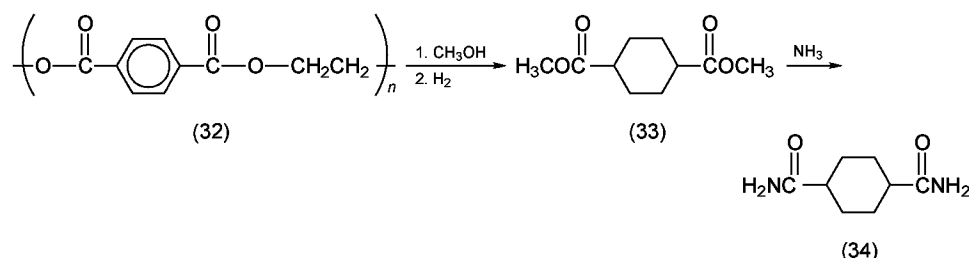
An alternative approach, generally referred to as a two-phase phosgenation, has gained wide scale acceptance for the production of aliphatic isocyanates (82,83). Typically, a cooled solution of phosgene and amine are mixed in the presence of concentrated sodium hydroxide [1310-73-2]. Reaction rates are very fast and overall product yields exceed 90%. Refinements in the reaction conditions have reduced the by-product formation arising from the reaction of the hydroxide with phosgene to less than 10%. This approach allows for the preparation of isocyanates containing labile groups, such as alkoxy (RO), which would be lost in a traditional high temperature, amine–hydrogen chloride salt phosgenation (39).

Low boiling isocyanates, such as methyl isocyanate [624-83-9], are difficult to prepare via conventional phosgenation due to the fact that the *N*-alkyl carbamoyl chlorides are volatile below their decomposition point. Interestingly, *N*-ethyl carbamoyl chloride decomposes at its boiling point whereas the *N*-propyl carbamoyl chloride is thermolyzed cleanly into isocyanate and hydrogen chloride.

A convenient method for the synthesis of these low boiling materials consists of the reaction of *N,N'*-dimethylurea [96-31-1] with toluene diisocyanate to yield an aliphatic–aromatic urea (84). Alternatively, an appropriate aliphatic–aromatic urea can be prepared by the reaction of diphenylcarbamoyl chloride [83-01-2] with methylamine. Thermolysis of either of the mixed ureas produces methyl isocyanate in high yield (3,85).

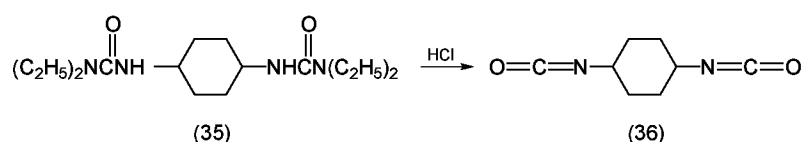


Akzo has been instrumental in developing a new process for the stereospecific synthesis of *trans*-1,4-cyclohexane diisocyanate [7517-76-2] (21). This process, based on the conversion of poly(ethylene terephthalate) [25038-59-9], circumvents the elaborate fractional crystallization procedures required for the existing *p*-phenylenediamine [108-45-2] approaches. The synthesis starts with poly(ethylene terephthalate) (PET) (32) or phthalic acid, which is converted to the dimethyl ester and hydrogenated to yield the cyclohexane-based diester (33). Subsequent reaction of the ester with ammonia provides the desired bisamide (34). The synthesis of the amide is the key



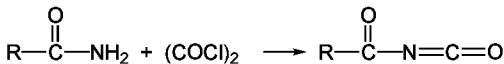
to the selectivity of this route. Typically, dimethyl 1,4-cyclohexane dicarboxylate [94-60-0] (33) is dissolved in a solvent. Ammonia is introduced. Methyl alcohol is removed by cracking and distillation. The desirable *trans*-amide is noted to precipitate upon prolonged reflux under an ammonia atmosphere.

Subsequent chlorination of the amide takes place in a two-phase reaction mixture (a dispersion of diamide in hydrochloric acid) through which a chlorine stream is passed. The temperature of this step must be maintained below 10°C to retard the formation of the product resulting from the Hofmann degradation of amides. Reaction of the *N,N*-dichloroamide with diethylamine [109-89-7] in the presence of base yields *trans*-1,4-cyclohexane-bis-1,3-diethylurea (35), which is transformed to the urea hydrochloride and pyrolyzed to yield the diisocyanate (36).

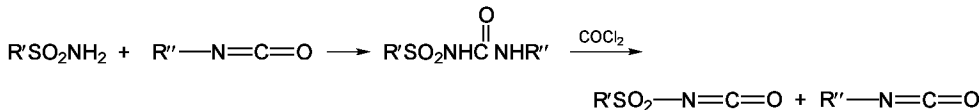


More recently, other nonphosgene routes for the preparation of aliphatic isocyanates have been reported. For example, American Cyanamid has disclosed the reaction of diisopropenylbenzene with HCl and isocyanic acid [75-13-8] to yield tetramethylxylylene diisocyanates (57). Monsanto has disclosed the use of carbon dioxide-amine complexes which are dehydrated, at low temperatures, with phosphoryl chloride [10025-87-3] or thionyl chloride [7719-09-7], as a viable route to a variety of aliphatic isocyanates. The process relies on the facile formation of the intermediate salt (30).REPLACEVariations of this process, in which phosgene is used as a dehydrating agent, have been reported earlier (84). Table 2 lists commercially available aliphatic isocyanates.

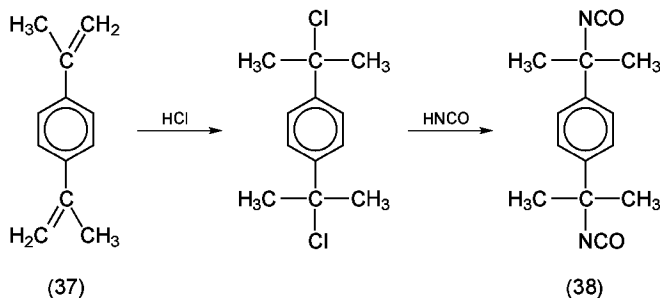
Specialty Isocyanates. Acyl isocyanates, extensively used in synthetic applications, cannot be directly synthesized from amides and phosgene. Reactions of acid halides with cyanates have been suggested. However, the dominant commercial process utilizes the reaction of carboxamides with oxalyl chloride [79-37-8]. Cyclic intermediates have been observed in these reactions which generally give a high yield of the desired products (86).



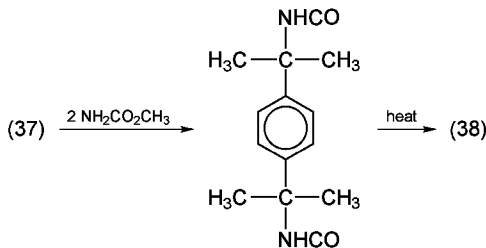
Commercially important arenesulfonyl isocyanates are not directly accessible from the corresponding sulfonamides via phosgenation due to lack of reactivity or by-product formation at elevated temperatures. A convenient method for their preparation consists of the reaction of alkyl isocyanates with sulfonamides to produce mixed ureas which, upon phosgenation, yield a mixture of alkyl and arenesulfonyl isocyanates. The desired product can be obtained by simple distillation (16). Optionally, the oxalyl chloride route has been employed for the synthesis of arenesulfonyl isocyanate (87).



Of the many other methods leading to isocyanates, only a few are practical enough in regard to availability of starting materials to be of general applicability. One of the more promising approaches utilizes olefinic substrates which add isocyanic acid in Markovnikov fashion to form alkyl isocyanates (2,9,87). This reaction is used to produce 1,4-bis(2-isocyanatoisopropyl)benzene (38) from cumene [98-82-8] (37) in commercial quantities (22,57,88). One approach uses the slow addition of the olefin to an excess of solvent and isocyanic acid in the presence of a catalytic amount of inorganic acid (57,88). Reaction temperatures are preferably maintained between 25 and 80°C. In the case of a diolefin, such as diisopropenylbenzene, the reaction can be controlled to favor the production of either the mono- or the diisocyanate by controlling the stoichiometry of the isocyanic acid in the reaction mixture. Another approach involves the formation of the dichloro intermediate. The dichloro compound reacts at low temperatures with an excess of isocyanic acid in the presence of a Lewis acid.

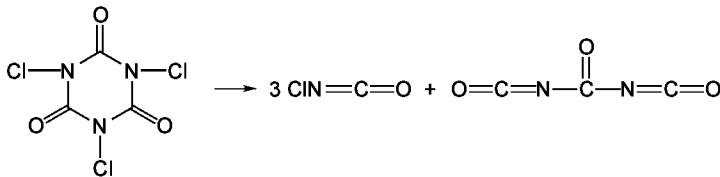


An alternative approach involves the reaction of an alkyl carbamate with a tertiary olefin (89,90). The resultant carbamates are thermally cracked at temperatures of 150–350°C to yield the isocyanate. The isocyanate is generally purified via distillation.

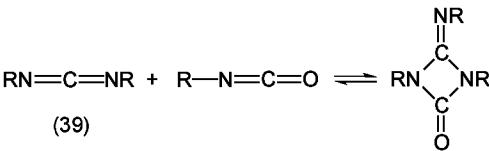


The exchange of halogen by isocyanato groups has also been suggested as a method of preparing isocyanates from chloro- or bromoalkanes (10). Metal cyanates are the reagents of choice for these exchange reactions which often entail the formation of oligomers of the desired isocyanates. For example, ethyl isocyanate [109-90-0] can be prepared in 90% yield by the reaction of ethyl bromide [74-96-4] with potassium isocyanate (91,92).REPLACEThe use of polar solvents, such as *N,N*-dimethylformamide [68-12-2], is noted to result in extensive trimer formation. However, if the isocyanate is trapped using compounds such as alcohols, carboxylic acids, and amines which contain active hydrogen, high yields are obtained (93).

Pyrolysis approaches can also be used to prepare substituted isocyanates which cannot be prepared using other methods. For example, *N,N,N'*-trichlorocyanuric acid [87-90-1] thermally dissociates to yield chloroisocyanate [13858-09-8] and carbonyl diisocyanate [6498-10-8]. The carbonyl isocyanate is unstable and polymerizes (8,94). Table 3 lists specialty isocyanates.



Carbodiimide Formation. Carbodiimide formation has commercial significance in the manufacture of liquid MDI. Heating of MDI in the presence of catalytic amounts of phosphine oxides or alkyl phosphates leads to partial conversion of isocyanate into carbodiimide (95). The carbodiimide (39) species reacts with excess isocyanate to form a 2 + 2 cycloaddition product. The presence of this product in MDI leads to a melting point depression and thus a mixture which is liquid at room temperature.



Analytical Methods

Organic isocyanates are generally characterized with respect to NCO content. A number of spectroscopic and chemical methods for isocyanate determination are available (96).

Isocyanates can be characterized using a strong absorption at 2300–2200 cm^{−1} in the ir spectrum. The position of the absorbance is influenced by conjugation and neighboring polar groups. This method has been successfully used in both kinetic and post-mortem characterizations of many polyurethane polymers.

Titration with dibutylamine [111-92-2] can also be used to determine the NCO content of isocyanates and prepolymers. Generally, an excess of amine in a suitable solvent such as chlorobenzene [108-90-7] is added to the sample. The resulting solution is allowed to react and the unreacted amine is back-titrated with dilute hydrochloric acid. For low NCO content levels, a colorimetric method is often used. The isocyanate-containing species is titrated with amine and the unreacted amine is determined using malachite green [569-64-2].

Both vapor-phase chromatography and high performance liquid chromatography, along with nuclear magnetic resonance spectroscopy, have been used for isomer and composition analysis.

Health and Safety Factors

Isocyanates are classified as dangerous substances (EEC Guidelines). They are generally labeled toxic and should be handled with care (97). Exposure hazards increase substantially when handling vapors or mists. Isocyanate vapors or mists may be irritating to the nose, throat, and lungs. Even brief exposure may cause irritation, difficult breathing, or coughing. Sensitization may result from excessive exposure. Subsequent exposure to low concentrations has been known to provoke allergic reactions with asthma-type symptoms. Industrially, this is important because MDI-based isocyanates, although having a low vapor pressure, can become airborne during machine flushing and filling procedures. Conversely, toluene diisocyanate (TDI) has a relatively high vapor pressure at ambient temperatures and a vapor density six times that of air. Thus open containers have the potential to yield high concentrations of TDI vapor. Also, many aliphatic isocyanates have high vapor pressures and must, therefore, be handled with special caution. Inhalation of aliphatic isocyanates is reported to retard the growth of laboratory mice (98). A compilation of toxicity information for several of the more common commercial isocyanates is presented in Table 4.

Table 4. Acute Toxicity of Diisocyanates in Rats

Isocyanate	LC ₅₀ (rat oral), g/kg		LC ₅₀ (rat inhalation) ^a , ppm	Concentration in air, ppm (STP)
TDI	5,800		110	19.6
MDI	> 31,600		370	0.1
IPDI	> 2,500		123	0.34
HDI	0.35–1		44–370	
PDI		0.9		
NDI ^b	> 10,000			c

^a Aerosol, 4 h.

^b Naphthalene-1,5-diisocyanate.

^c Negligible vapor pressure at 50°C.

Repeated or prolonged skin contact may cause irritation, blistering, dermatitis, or skin sensitization. Contact with the eye has been reported to cause irritations in testing with rabbits. For these reasons, isocyanates must be handled in well-ventilated areas. Respirators should be worn whenever the possibility of vapor exposure exists. If inhalation occurs, the affected person should be moved to a well-ventilated area. Chemical goggles should be worn when handling isocyanates. All work areas should be equipped with an eye wash. In the event of eye contact, the eye should be irrigated immediately. The eyes should be held open while flushing with a continuous low pressure stream of water for at least 15 minutes. In the event of direct skin contact, use a safety shower immediately, removing all clothing while washing. In all cases, call a physician immediately.

The most overlooked hazard and contaminant is water (99). Water reacts with isocyanates at room temperature to yield both ureas and large quantities of carbon dioxide. The presence of water or moisture can produce a sufficient amount of CO₂ to overpressurize and rupture containers. As little as 30 mL of water can result in 40 L of carbon dioxide which could result in pressures of up to 300 kPa (40 psi). For these reasons, the use of dry nitrogen atmospheres is recommended during handling. If a plant air system must be used, purification equipment, such as oil traps and drying beds, should be installed between the source and the isocyanate vessel.

Also, the presence of strong bases, even in trace amounts, can promote the formation of isocyanurates or carbodiimides. In the event of gross contamination, the exothermic reaction can sharply increase the temperature of the material. Normally, the trimerization reaction occurs first and furnishes heat for the carbodiimide reaction. The carbodiimide reaction liberates carbon dioxide and forms a hard solid. The liberation of carbon dioxide in a sealed vessel could result in overpressurization and rupture.

Temperature control is important in the handling and storage of isocyanates. Storage at inappropriate temperatures can cause product discoloration, viscosity increases, and dimerization. Handling personnel should consult the technical data sheets for the recommended storage temperature of the specific isocyanate product.

Most commercial isocyanates have a high flash point and are classified as Class IIIB combustible liquids. These materials, however, burn in the presence of an existing fire or heat source in the presence of oxygen. In the event of an isocyanate fire, use a carbon dioxide or dry chemical extinguisher. For fires covering large areas, use of a protein foam or water spray is recommended. Personnel engaged in fighting isocyanate fires must be protected against nitrogen dioxide vapors and isocyanate fumes. Firefighters should wear approved positive pressure, self-contained breathing apparatus and fire-resistant clothing.

Economic Aspects and Applications

Since 1971, the overall demand for isocyanates has increased at a compounded rate of 12%. Although this level will not likely be sustained in the future due to the maturation of key application markets, it is probable that additional growth will occur through the year 2000. This trend will likely include a shift in emphasis from TDI to MDI and polymeric MDI-based materials. New growth opportunities in the construction industry, structural applications, and growth in the automotive industry exist. Third-world markets are also anticipated to provide growth opportunities.

Globally, BASF, Bayer (Miles in North America), Dow, and ICI historically have been the leading producers of aromatic isocyanates. In North America, Olin is a principal supplier of TDI and aliphatic isocyanates. Rhône-Poulenc and Hoechst are principal suppliers in Europe. A listing of all the principal global suppliers and their respective products and trade names is presented in Tables 5 and 6. A breakdown of isocyanate demand by region is presented in Table 7.

Table 5. 1993 Isocyanate Product Capacities, 10³ t

Region	TDI	MDI	PMDI	Aliphatic
North America				
BASF	95	12.5	100	
Dow	70	40.5	163.5	
ICI (Rubicon)	32.5	25	150	
Miles	132.5	35	167.5	16
Olin	110			
Total	485	113	592.5	24.5
Europe				
ICI		11.5	142.5	
BASF	17.5	29.5	173	pilot
Bayer	206	67.5	298.5	10
Montedipe	121	3.5	88	
Rhône-Poulenc	126.5			5
Dow		39.5	181.5	
Total	530	176.7	905	27.5
Pacific				
Korea Fine Chemical	27.5			
Mitsui Toatsu	61		60.5	0.2
Takeda	82.5			0.6
BASF		12.5	30	
Dow Mitsubishi		14	15.5	
Nippon Polyurethanes		20	110	
Sumitomo Bayer		12.5	41	
Sunghua		11.5	27	
Total	308	77.5	306	3.3
Latin America				
Pronor	70.5		22	
Bayer	13		22	
Total	102.5	7.5	44	

Table 6. Trademarks for Global Isocyanate Manufacturers

Manufacturer	Trademark	Region
Akzo	Elate	North America
Bayer	Desmodur	Europe
BASF	Ekanate	North America
BASF	Lupranat	Europe
BASF-Schwarzheide	Systanat	Europe
Cariosa	Cortume	Europe
Dainippon	Sothanate	Pacific
Dainichi Saika	Resamine	Pacific
Dow	Isonate	North America
Dow	Voranate	North America
ICI	Suprasec	Europe
Interchem	Prepol	Europe
Lancro	Quasilan	Europe
Lancro	Isocon	Europe
Miles	Multrathane	North America
Miles	Mondur	North America
Montedison	Tedimon	Europe
Olin	Olin-TDI	North America
Polymer Chemicals	Polidur	Europe
Reichold	Polylite	North America
Rheinchemie	Rhenodur	Europe
Rhône-Poulenc	Scurane	Europe
Rhône-Poulenc	Tolonate	Europe
Rubicon	Rubinate	North America
Shell	Caradate	Europe
Takeda	Takenat	Europe
Thanex	Poronat	Europe
Thiokol	Solithane	North America

Table 7. 1990 Regional Demand for Polyurethane Polymers, 10³ t

Item	North America ^a	Europe	Pacific
flexible foams furniture bedding transportation	940	1127	510
other			
rigid foams appliance building construction	450	405	177

elastomers ^b automotive RIM footwear	143	111	102
elastomers coatings adhesives ^c	234	268	219
other ^d	106	183	148
<i>Total</i>	<i>1873</i>	<i>2094</i>	<i>1156</i>

^a Includes Latin America.
^b Includes thermoplastic urethanes and synthetic leather.
^c Includes paints.
^d Includes fibers, marine flotation, film, and textiles.

Aromatic Isocyanates. In North America, aromatic isocyanates are heavily used as monomers for addition and condensation polymers. The principal applications include both flexible and rigid polyurethane foam and noncellular applications, such as coatings, adhesives, elastomers, and fibers.

In the North American flexible foam market, toluene diisocyanate, an 80:20 mixture of the 2,4- and 2,6-isomers, is the monomer of choice due to its low cost and general availability. In both Europe and Japan, MDI blends and prepolymers are finding increased usage. Prepolymers are an adduct or reaction intermediate of a polyol and a monomeric isocyanate. The reaction of the isocyanate and a polyalkylene oxide polyol is the primary reaction which forms the matrix. Density reduction is provided via the *in situ* generation of carbon dioxide which arises from the simultaneous water–isocyanate reaction. Optionally, graft copolymers and auxiliary blowing agents can be used. Flexible foams can be produced using either a continuous or molded procedure. The products from the continuous process (buns) are cut to shape for the particular application, while the molded foams are formed to shape during the molding process. The density of the continuous buns is traditionally between 16–32 kg m^{−3} (1–2 lbs ft^{−3}). Molded foams have densities greater than 32 kg m^{−3}. Flexible foams find applications as automotive cushions, carpet underlay, furniture, seating, and bedding.

Rigid foams are based primarily on polyfunctional, low molecular weight alcohols and amines. Most global applications conventionally use polymeric isocyanates, TDI, or an undistilled grade of mixed TDI isomers. TDI prepolymers which have hydroxyl and isocyanate groups have been marketed as a low vapor pressure alternative to undistilled TDI. Density reduction is effected via the addition of chlorofluorocarbons, low molecular weight alkanes, or via the *in situ* generation of carbon dioxide. The resultant closed cell foams find applications as insulators in construction, appliance, transportation, pipeline, and tank end uses.

Similarly, polyisocyanurate (PIR) rigid foams are based on PMDI and polyester or polyether polyols (and blends thereof). The cross-link density, aromatic content, cell wall thickness, and polymer distribution are known to have a pronounced effect on the overall performance of the foam. Polyisocyanurate foams are used extensively in industrial applications having service temperature ranges from −200 to 150°C. These foams are known to provide efficient thermal insulation characteristics and structural integrity to a variety of composite applications. Apart from roofing and sheathing, PIR foams are used in garage doors, building panels, and foaming applications. They can also be supplied in the form of boardstock for fabrication into sheets, pipe covering, and other shapes.

Adhesives and coatings formulations utilize both MDI and TDI along with a variety of polyether and polyester polyols. The largest segment of the business is the one-part or moisture-cured approach, which is heavily reliant on prepolymers. Optionally, blocked isocyanate technology may be used to provide systems having an extremely long shelf life. Suitable blocking agents include phenols, diethyl malonate [105-53-3], acetone oxime [127-06-0], and ε-caprolactam [105-60-2]. Typical applications include flexible film packaging and wood furniture assembly. Two-part reactive adhesives employ a mixture of pure isocyanates and prepolymers and are primarily used for industrial product assembly and commercial construction.

Elastomers are segmented block copolymers. They employ a mixture of pure MDI-based isocyanates and prepolymers. The hard blocks consist of high melting MDI–glycol units which aggregate via hydrogen bonding to provide a high degree of virtual cross-linking. Suitable glycols include ethylene glycol [107-21-1], butanediols, and hexanediols. The soft block consists of high molecular weight poly(alkylene oxide) polyols. These materials are noted for their resiliency, abrasion resistance, solvent resistance, and the high level of tensile, tear, and elongation performance. Applications include shoe soles, wheels, rollers, belts, gaskets, and thermoplastic polyurethanes. Methylene diphenyl diisocyanate (MDI) is used extensively in the production of Spandex fibers. Spandex is used in foundation garments and swimwear (see FIBERS, ELASTOMERIC). Typical demand for various applications is presented in Table 7.

Aliphatic Isocyanates. Aliphatic diisocyanates have traditionally commanded a premium price because the aliphatic amine precursors are more expensive than aromatic diamines. They are most commonly used in applications which support the added cost or where the long-term performance of aromatic isocyanates is unacceptable. Monofunctional aliphatic isocyanates, such as methyl and *n*-butyl isocyanate, are used as intermediates in the production of carbamate-based and urea-based insecticides and fungicides (see FUNGICIDES, AGRICULTURAL; INSECT CONTROL TECHNOLOGY).

A number of markets have been established for light-stable, aliphatic diisocyanates in the United States. The largest market is in high performance coatings (see COATINGS). The largest coatings market is in automotive refinishes. Other coatings include uv-cured coatings for vinyl tile and sheet flooring, electronic circuit boards, powder coatings, and paints. Hydrogenated MDI (H₁₂MDI), *m*-xylylene diisocyanate (XDI), and isophorone diisocyanate [4098-71-9] are currently used in many of these coating applications.

Aliphatic isocyanates have a small but growing market application in thermoplastic polyurethanes (TPU). Medical applications include wound dressings, catheters, implant devices, and blood bags. A security glass system using light-stable TPU as an inner layer is under evaluation for shatterproof automotive windshield applications.

Developments in aliphatic isocyanates include the synthesis of polymeric aliphatic isocyanates and masked or blocked diisocyanates for applications in which volatility or reactivity are of concern. Polymeric aliphatic isocyanates are made by copolymerizing methacrylic acid derivatives, such as 2-isocyanatoethyl methacrylate, and styrene [100-42-5] (100). Blocked isocyanates are prepared via the reaction of the isocyanate with an active hydrogen compound, such as ε-caprolactam, phenol [108-95-2], or acetone oxime.

Specialty Isocyanates. Specialty isocyanates are organic isocyanates having the isocyanate function attached to a carbonyl group or to elements other than carbon. This group is appropriately named because they find use in highly specialized applications. *p*-Toluenesulfonyl isocyanate [4083-64-1] is used as a drying agent for organic solvents. Arenesulfonyl diisocyanates, such as *m*-phenylenedisulfonyl diisocyanate, are used as monomers for base-soluble polymers. Arenesulfonyl monoisocyanates are used as intermediates for pharmaceuticals and herbicides (see HERBICIDES; PHARMACEUTICALS).

BIBLIOGRAPHY

"Isocyanates, Organic" in *ECT* 2nd ed., Vol. 12, pp. 45–64, by D. H. Chadwick, Mobay Chemical Co., and E. E. Hardy, Monsanto Research Corp.; in *ECT* 3rd ed., Vol. 13, pp. 789–818, by D. H. Chadwick and T. H. Cleveland, Mobay Chemical Co.

1. A. A. A. Sayigh, H. Ulrich, and W. J. Farrissey, in J. K. Stilles and T. W. Campbell, eds., *Condensation Monomers*, John Wiley & Sons, Inc., New York, 1972, pp. 369–476.
2. R. Richter and H. Ulrich, in S. Patai, ed., *The Chemistry of Cyanates and their Thio Derivatives*, John Wiley & Sons, Inc., New York, 1977, p. 619.
3. W. Siefken, *Liebigs Ann. Chem.* **17**, 1284 (1949).
4. J. K. Rassmussen and A. Hassner, *Chem. Rev.* **76**, 389 (1976).
5. H. J. Twitchett, *Chem. Soc. Rev.* **3**, 209 (1974).
6. H. Ulrich, *J. Elastoplastics* **3**, 97 (1971).
7. V. I. Gorbartenko and L. I. Samarai, *Synthesis* **1980**, 85 (1980).
8. H. Hagemann, *Angew. Chem.* **89**, 789 (1977).
9. H. Hagemann, *Houben-Weyl: Kohlensaure-Derivate*, G. Thieme, New York, 1983.

10. A. Wurtz, *Compt. Rend.* **27**, 242 (1848).
11. H. Ulrich and R. Richter, in W. Forest, ed., *Newer Methods of Preparative Organic Chemistry-VI*, Academic Press, Inc., New York, 1970, p. 267.
12. W. Hentschel, *Chem. Ber.* **17**, 1284 (1884).
13. H. Staudinger and E. Hauser, *Helv. Chim. Acta.* **4**, 861 (1921).
14. D. N. Dhar and K. S. K. Murthy, *Synthesis* **1986**, 437 (1986).
15. H. Ulrich, R. Richter, and B. Tucker, *Synthesis* **1979**, 277 (1979).
16. H. Ulrich, B. Tucker, and A. A. R. Sayigh, *J. Org. Chem.* **31**, 2658 (1966).
17. H. Ulrich, *Chem. Rev.* **65**, 869 (1965).
18. P. A. S. Smith, *Organic Reactions* **3**, 337 (1946).
19. F. L. Pyman, *J. Chem. Soc.* **103**, 852 (1913).
20. H. Ulrich and R. Richter, *J. Org. Chem.* **38**, 2557 (1973).
21. **U. S. Pat. 4,223,145** (1979) P. Hentschel, H. Zengel, and M. Bergfeld (to Akzo).
22. K. A. Henderson and V. A. Alexandian, *Org. Prep. Proc. Int.* **18**, 149 (1986).
23. **Ger. Pat. 42,343,826** (1974) J. A. Zajacek, J. J. McCoy, and K. E. Fuger (to Arco).
24. **U.S. Pat. 3,895,054** (1974) J. A. Zajacek, J. J. McCoy, and K. E. Fuger (to Arco).
25. **U.S. Pat. 4,369,141** (1974) J. A. Zajacek, J. J. McCoy, and K. E. Fuger (to Arco).
26. **Ger. Pat. 52,635,490** (1977) R. Tsumura, U. Takaki, and A. Takeshi (to Mitsui-Toatsu).
27. **Ger. Pat. 123,807** (1975) K. Schwetlick and co-workers (to VEB Schwarzeide).
28. **U.S. Pat. 4,268,683** (1981) A. E. Gurgiolo (to Dow Chemical).
29. **U.S. Pat. 3,366,662** (1968) E. H. Kober and E. Smith (to Olin Chemical).
30. **U.S. Pat. 5,189,205** (1993) W. D. McGhee and T. E. Waldman (to Monsanto).
31. **U.S. Pat. 3,960,914** (1992) J. E. Lyons (to Sun Ventures, Inc.).
32. B. Hileman, *Chem. Eng. News* **70**(45), 9 (1992).
33. L. Capuano and co-workers, *Chem. Ber.* **106**, 312 (1973).
34. H. Brederick, G. Simchen, and E. Goknel, *Chem. Ber.* **103**, 236 (1970).
35. R. Richter, *Chem. Ber.* **105**, 82 (1972).
36. M. Coenen, *Chem. Ber.* **80**, 546 (1947).
37. F. Effenberger and co-workers, *Chem. Ber.* **101**, 502 (1968).
38. W. J. Kaufmann, *J. Org. Chem.* **39**, 2472 (1974).
39. **Ger. Pat. 947,471** (1953) J. Pfirschke (to Bayer); J. E. Herweh and W. J. Kaufmann, *J. Heterocycl. Chem.* **8**, 983 (1971).
40. M. E. Dyen and D. Swern, *Chem. Rev.* **67**, 197 (1967).
41. **Ger. Pat. 1,909,573** (1970) K. H. Markiewicz (to Atlas Chemical Industries, Inc.).
42. H. V. Brachel and R. Merten, *Angew. Chem.* **74**, 872 (1962).
43. H. Biener, *Liebigs Ann. Chem.* **686**, 102 (1965).
44. H. Bohme and W. Pasche, *Arch. Pharm.* **302**, 81 (1969).
45. H. Ulrich and A. A. R. Sayigh, *Angew. Chem. Int. Ed.* **5**, 704 (1966).
46. J. Goerdeler and H. Schenic, *Chem. Ber.* **98**, 2934, 3831 (1965).
47. J. L. Chitwood, P. J. Gott, and J. C. Martin, *J. Org. Chem.* **36**, 2228 (1971).
48. R. J. P. Barends, W. N. Speckamp, and H. O. Huisman, *Tetrahedron Lett.* **1970**, 5301 (1970).
49. W. Seeliger and co-workers, *Angew. Chem.* **78**, 913 (1966).
50. R. Gompper, *Angew. Chem.* **81**, 348 (1969) and references therein.
51. E. Van Look, *Ind. Chem. Belg.* **1974**, 661 (1974).
52. G. Denecker, G. Smets, and G. L'Abbe, *Tetrahedron* **31**, 765 (1975).
53. G. M. Vandersavel, G. Smets, and G. L'Abbe, *J. Org. Chem.* **38**, 675 (1973).
54. R. B. Wilson and co-workers, *J. Am. Chem. Soc.* **105**, 1672 (1983).
55. R. Richter and H. Ulrich, *Synthesis* **1975**, 463 (1975).
56. A. Etienne, B. Bonte, and B. Druet, *Bull. Soc. Chim. Fr.* **1972**, 251 (1972).
57. **U. S. Pat. 4,377,530** (1983) S. L. Trenbeath, A. M. Feldman, and L. J. Nummy (to American Cyanamid).
58. O. K. White and F. D. Greene, *J. Org. Chem.* **43**, 4530 (1978).
59. G. Ostrogovich and co-workers, *Tetrahedron* **25**, 1875 (1969).
60. **U. S. Pat. 3,056,835** (1962) J. J. Monagle and H. R. Nace (to Du Pont).
61. H. Ulrich, in W. F. Gum, W. Riese, and H. Ulrich, eds., *Reaction Polymers*, Hanser, New York, 1992, p. 358.
62. K. H. Slotta and T. Tschesche, *Chem. Ber.* **60**, 265 (1927).
63. V. E. Shashoua, *J. Am. Chem. Soc.* **81**, 3156 (1959). V. E. Shashoua, W. Sweeny, and R. F. Tietz, *J. Am. Chem. Soc.* **82**, 866 (1960).
64. R. C. Schulz and M. Hartmann, *Makromol. Chem.* **55**, 227 (1962).
65. D. W. Baugh, Jr., in Ref. 61, p. 259.
66. G. Oertel, *Polyurethane Handbook*, Hauser, New York, 1985.
67. G. Woods, *The ICI Polyurethanes Book*, John Wiley & Sons, Inc., New York, 1987.
68. S. Peterson, *Liebigs Ann. Chem.* **562**, 205 (1949).
69. R. D. Priester, Jr., N. P. Vespoli, and F. Martinez, III, in Ref. 61, p. 388.
70. R. L. Zapp, G. E. Sernivk, and K. S. Minckler, *Rubber Chem. Techn.* **43**, 1154 (1970).
71. S. Motoki, T. Saito, and H. Kagami, *Bull. Chem. Soc. Jpn.* **47**, 775 (1974).
72. P. Babusiaux, R. Longeray, and J. Dreux, *Liebigs Ann. Chem.* **3**, 487 (1976).
73. P. S. Carleton and W. J. Farrissey, *Tetrahedron Lett.* **1969**, 3485 (1969).
74. K. Onder, in Ref. 61, p. 405.
75. H. Ulrich, J. N. Tilley, and A. A. R. Sayigh, *J. Org. Chem.* **29**, 2401 (1964).
76. **U.S. Pat. 4,207,212** (1980) G. P. Balabanov and co-workers.
77. K. D. Kim and co-workers, *J. Mol. Catalysis* **75**, L1 (1992).
78. T. W. Leung and B. D. Dombek, *J. Chem. Soc., Chem. Comm.* **1992**, 205 (1992).
79. Y. Izumi and co-workers, *J. Mol. Catalysis* **72**, 37 (1972).
80. W. B. Hardy and R. P. Bennett, *Tetrahedron Lett.* **1967**, 961 (1967).
81. **Ger. Pat. 2,819,826** (1978) R. Becker and co-workers (to Bayer).
82. H. Ulrich, in Ref. 61, pp. 50–64.
83. **Ger. Pat. 1,668,109** (1968) D. Adt (to Bayer).
84. **Ger. Pat. 227,004** (1966) H. G. Henning (to Bayer).
85. W. Bunge, *Angew. Chem.* **72**, 1002 (1960).
86. A. J. Speziale and L. R. Smith, *J. Org. Chem.* **27**, 3742 (1962).
87. F. W. Hoover, H. S. Rothrock, and K. E. Olson, *J. Org. Chem.* **29**, 143 (1964).

88. U.S. Pat. **4,361,518** (1982) B. Singh and W. A. Henderson (to American Cyanamide).
89. Belg. Pat. 625,748 (1963) G. Mueller and R. Merton (to Bayer).
90. U.S. Pat. **4,439,616** (1984) B. Singh, L. W. Chang, and P. S. Forgione (to American Cyanamide).
91. U.S. Pat. **3,017,420** (1962) W. D. Schaeffer (to Union Oil of California).
92. U.S. Pat. **4,130,577** (1978) N. Nagato and T. Naito (to Showa Denko).
93. Brit. Pat. **858,810** (1961) K. Fukui and H. Kitano.
94. Ger. Pat. **2,408,069** (1975) H. Hagemann (to Bayer).
95. Brit. Pat. **1,069,858** (1967) W. J. Farrissey and A. Odinak (to Upjohn).
96. L. B. Coyne, in Ref. 61, p. 794.
97. M. J. Blankenship, in Ref. 61, p. 788.
98. A. Ader, E. Loeser, and I. F. Larney, in *Proceedings of the Polyurethanes World Congress*, Aachen, Germany, Sept. 1987, and References therein.
99. M. J. Blankenship, in Ref. 61, p. 771.
100. M. R. Thomas, *Preprints from Division of Organic Coatings and Plastic Chemistry, 183rd National ACS Meeting*, Las Vegas, 1982, p. 506.

Reinhard H. Richter
Ralph D. Priester, Jr.
Dow Chemical

ISOCYANURIC COMPOUNDS.

See CYANURIC AND ISOCYANURIC ACIDS.

ISOPHORONE.

See KETONES.

ISOPRENE

From the time that isoprene was isolated from the pyrolysis products of natural rubber (1), scientific researchers have been attempting to reverse the process. In 1879, Bouchardat prepared a synthetic rubbery product by treating isoprene with hydrochloric acid (2). It was not until 1954–1955 that methods were found to prepare a high *cis*-polyisoprene which duplicates the structure of natural rubber. In one method (3,4) a Ziegler-type catalyst of trialkylaluminum and titanium tetrachloride was used to polymerize isoprene in an air-free, moisture-free hydrocarbon solvent to an all *cis*-1,4-polyisoprene. A polyisoprene with 90% 1,4-units was synthesized with lithium catalysts as early as 1949 (5). With the availability of polymerization catalysts, extensive efforts were devoted to developing economical processes for manufacture of isoprene. Several synthetic routes have been commercialized. With natural rubber as an alternative, the ultimate value of the polymer was more or less dictated by that market. The first commercial use of isoprene in the United States started in 1940. It was used as a minor comonomer with isobutylene for the preparation of butyl rubber. Polyisoprene was commercialized extensively in the 1960s (6). In the 1990s isoprene is used almost exclusively as a monomer for polymerization (see ELASTOMERS, SYNTHETIC POLYISOPRENE). The isoprene unit exists extensively in nature. It is found in terpenes, camphors, diterpenes (eg, abietic acid), vitamins A and K, chlorophyll, and other compounds isolated from animal and plant materials. The correct structural formula for isoprene was first proposed in 1884 (7).

Properties

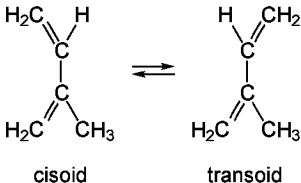
Isoprene [78-79-5] (2-methyl-1,3-butadiene) is a colorless, volatile liquid that is soluble in most hydrocarbons but is practically insoluble in water. Isoprene forms binary azeotropes with water, methanol, methylamine, acetonitrile, methyl formate, bromoethane, ethyl alcohol, methyl sulfide, acetone, propylene oxide, ethyl formate, isopropyl nitrate, methylal (dimethoxymethane), ethyl ether, and *n*-pentane. Ternary azeotropes form with water–acetone, water–acetonitrile, and methyl formate–ethyl bromide (8). Typical properties of isoprene are listed in Table 1.

Table 1. Properties of Isoprene

Property	Value
mol wt	68.11
density of liquid, gm cm ⁻³ at 25°C	0.6759
freezing point, °C	145.95
bp at 101.3 kPa, °C	34.067
<i>n</i> _D ²⁰	1.41524
flash point, °C	48
autoignition, °C	220
dipole moment of liquid, Cm ^b	9.4 × 10 ⁻³¹
heat of combustion at 25°C, kJ mol ^c	146
heat of formation at 25°C, kJ mol ^c	
liquid	49.4
gas	75.78
coefficient of expansion (–20.6 to 21.1 °C)	0.0016

^a To convert kPa to atm, divide by 101.3.
^b To convert Cm to debye, multiply by 2.997 × 10²⁹.
^c To convert kJ to kcal, divide by 4.184.

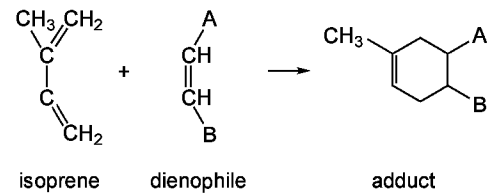
Many of the common properties of isoprene have been presented graphically (9). These include vapor pressure, heat of vaporization, liquid heat capacity, vapor heat capacity, liquid density, vapor viscosity, liquid viscosity, surface tension, and vapor thermal conductivity. **Conformation.** The exact conformation of the isoprene molecule is still in doubt. It is generally accepted that rotation is restricted around the central C—C single bond. Isoprene may be considered as an equilibrium of two conformations, namely a cisoid (*s-cis*) conformation in which both vinyl groups are located on the same side of the C—C bond, and a transoid (*s-trans*) one with the vinyl groups located on the opposite sides of the bond. The predominance of the trans-planar or nonplanar configuration has been supported by experimental data (10–14).

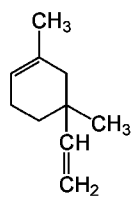


The energy required for a trans–cis change in conformation has been calculated to be 4.59 kJ mol^{–1} (1.10 kcal mol^{–1}) (14) and 6.28 kJ mol^{–1} (15). This difference is not considered sufficiently large to prevent easy interconversion (16), and is not great enough to provide a barrier to chemical reactions requiring the cis form, eg, in the Diels-Alder reaction of isoprene with maleic anhydride (14). The transoid geometry of butadiene is accepted as the more stable one in the ground state, but in the excited state, the cisoid and transoid forms are of nearly equal energy and of similar stability (17). Similar behavior is anticipated for isoprene. Conformational considerations are important in polymerization. Polymer structures with reference to the conformation of the monomer have been discussed (18–21).

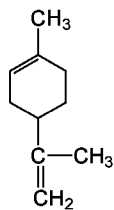
Reactions

Isoprene is highly reactive both as a diene and through its allylic hydrogens, and its reactions are similar to those of butadiene (qv) (8). Apart from polymerization, the most widely investigated isoprene reactions are the formation of six-membered rings by the Diels-Alder reaction:

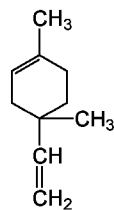




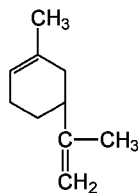
(3)
[1611-21-8]



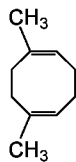
(4)
[138-86-3]



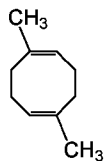
(5)
[1743-61-9]



(6)
[38738-60-2]



(7)
[3760-13-2]



(8)
[3760-14-3]

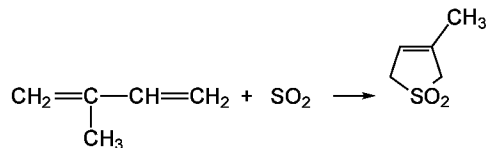
The proportions of the various dimers depend on the reaction conditions. Dimers (4) and (6) constitute 57–80% of the products that are formed from 100–190°C. At 520°C, compounds (3) and (4) are the principal reaction products (34). The rate of dimer formation as a function of temperature ranges from 0.000017% isoprene dimerized per hour at 20°C to 0.25% h at 100°C (35).

In the process of thermal dimerization at elevated temperatures, significant polymer is formed resulting in seriously decreased yields of dimer. Dinitroresol has been shown to be one of the few effective inhibitors of this thermal polymerization. In the processing of C₅ streams, thermal dimerization to convert 1,3-cyclopentadiene to dicyclopentadiene is a common step. Isoprene undergoes significant dimerization and codimerization under the process conditions.

The dimerization of isoprene has been accomplished by methods other than heating. Thus isoprene has been dimerized by uv radiation in the presence of photosensitizers to give a complex mixture of cyclobutane, cyclohexene, and cyclooctadiene derivatives (36,37). Sulfuric acid reportedly

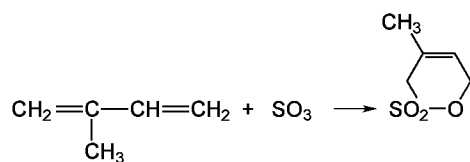
converts isoprene to linear and cyclic dimers (38). Ziegler-Natta coordination catalysts containing titanium (39,40), nickel (41), and iron (42–44) dimerize isoprene at low temperatures. In most instances, the proportions of dimers differ from those obtained in thermal dimerizations, and linear trimers and higher homologues are also obtained in the reaction products.

Isoprene can form five-membered hydrocarbon rings (8). A five-membered, sulfur-containing ring is an intermediate in an isoprene purification reaction:

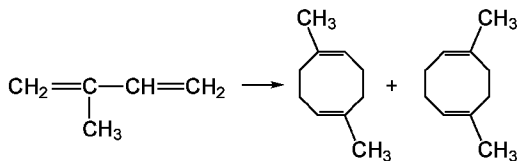


The solid sulfone can be recrystallized, and the isoprene can be regenerated by heating the purified sulfone (45,46).

With SO_3 –DMF (dimethylformamide) complex, a six-membered ring can form the delta sulfone (47):

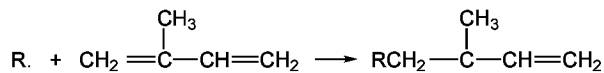


Eight-membered rings may be formed thermally or photochemically (see [PHOTOCHEMICAL TECHNOLOGY](#)). Excellent yields can be obtained with ferric acetylacetonate–triethylaluminum–dipyridyl (43).



Twelve-membered rings have been obtained using coordination catalysts. The *trans,trans,cis*-cyclododecatiene has been prepared with a tetrabutyl titanate–diethylaluminum chloride catalyst (48,49) and with a chromium-based system (50). The *trans,trans,trans*-isomer has been prepared with a nickel system.

Free-Radical Reactions. Free radicals attack isoprene, and two competing mechanisms, at the double bond or involving C–H bonds, are postulated:



The rates of these two reactions have been studied for the attack of trifluoromethyl (51) and methyl radicals (52) in isoprene that has been dissolved in 2,3-dimethylbutane and isooctane, respectively. The rate constants for the reactions with isoprene are much greater than those for the attack on the solvent. The ratio between the two rates for the attack of trifluoromethyl radicals varies from 1090 at 65°C to 233 at 180°C. For the corresponding reaction involving methyl radicals, the ratio is 2090 at 65°C.

The photosensitized dimerization of isoprene in the presence of benzil has been investigated. Mixtures of substituted cyclobutanes, cyclohexenes, and cyclooctadienes were formed and identified (53). The reaction is believed to proceed by formation of a reactive triplet intermediate. The energy for this triplet state presumably is obtained by interaction with the photoexcited benzil species. Under other conditions, photolysis results in the formation of a methylcyclobutene (54,55).

The addition of aromatic and aliphatic thiols, RSH and ArSH, and a thioacetic acid to isoprene yields mainly the *trans*-1,4-adduct (56). The aromatic thiyl radicals, ArS $\cdot$, add almost entirely to the first carbon atom; however, aliphatic thiyl radicals, RS $\cdot$, also add to the fourth C atom in significant amounts.

Halogens and Halogenated Compounds. The chlorination of isoprene in CCl_4 at -5 to -10°C , using an equimolar ratio of chlorine to isoprene, gives a mixture of 44% of 1,4-dichloro-2-methyl-2-butene and 14% of 3,4-dichloro-2-methyl-1-butene as addition products, along with 42% of the substitution product, 2-chloromethyl-1,3-butadiene (57). For the latter product, the reaction is an electrophilic substitution, and the carbon bearing the chlorine is not of the original methyl group in the isoprene molecule (58). At elevated temperatures (850°C) and in the vapor state, chlorine does not add but chlorinates the methyl in a free-radical substitution reaction. Thus a given reagent reacts by different mechanisms under different conditions.

Bromination of isoprene using Br_2 at -5°C in chloroform yields only *trans*-1,4-dibromo-2-methyl-2-butene (59). Dry hydrogen chloride reacts with one-third excess of isoprene at -15°C to form the 1,2-addition product, 2-chloro-2-methyl-3-butene (60). When an equimolar amount of HCl is used, the principal product is the 1,4-addition product, 1-chloro-3-methyl-2-butene (61). The mechanism of addition is essentially all 1,2 with a subsequent isomerization step which is catalyzed by HCl and is responsible for the formation of the 1,4-product (60). The 3,4-product, 3-bromo-2-methyl-1-butene, is obtained by the reaction of isoprene with 50% HBr in the presence of cuprous bromide (59). Isoprene reacts with the reactive halogen of 3-chlorocyclopentene (62).

The reaction of dihalocarbenes with isoprene yields exclusively the 1,2- (or 3,4-) addition product, eg, dichlorocarbene Cl_2C : and isoprene react to give 1,1-dichloro-2-methyl-2-vinylcyclopropane (63). The evidence for the presence of any 1,4 or much 3,4 addition is inconclusive (64). The cycloaddition reaction of 1,1-dichloro-2,2-difluoroethylene to isoprene yields 1,2- and 3,4-cycloaddition products in a ratio of 5.4:1 (65). The main product is 1,1-dichloro-2,2-difluoro-3-isopropenylcyclobutane, and the side product is 1,1-dichloro-2,2-difluoro-3-methyl-3-vinylcyclobutane. When the dichlorocarbene is generated from CHCl_3 , plus aqueous base with a tertiary amine as a phase-transfer catalyst, the addition has a high selectivity that increases (for a series of diolefins) with a decrease in activity (66) (see [CATALYSIS, PHASE TRANSFER](#)). For isoprene, both mono-(1,2-) and diadducts (1,2- and 3,4-) could be obtained in various ratios depending on which amine is used.

Isoprene reacts with α -chloroalkyl ethers in the presence of ZnCl_2 in diethyl ether from 0–10°C. For example, α -chloromethyl methyl ether at 10°C gives a 6:1 ratio of the 1,4-adduct, (*E*)-4-chloro-1-methoxy-2-methyl-2-butene, to the 1,2-adduct, 2-chloro-1-methoxy-2-methyl-3-butene. Other α -chloroalkyl ethers react in a similar manner to give predominately the 1,4-addition product. A wide variety of allylic chlorides and bromides and α -chloroethers and esters add primarily 1,4- to isoprene in the presence of acid catalysts (8).

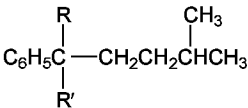
A telomerization reaction of isoprene can be carried out by treatment with 2-chloro-3-pentene, prepared by the addition of dry HCl to 1,3-pentadiene (67). An equimolar amount of isoprene in dichloromethane reacts with the 2-chloro-3-pentene at 10°C with stannic chloride as catalyst. 1-Chloro-3,5-dimethyl-2,6-octadiene is obtained in 80% yield by 1,4-addition.

Addition reactions between isoprene and tetrahalomethanes can be induced by peroxides, high energy ionizing radiation, or other radical-generating

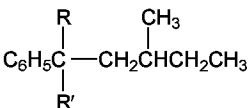
catalysts (see INITIATORS). In a radical reaction (68), carbon tetrachloride adds 1,4- to isoprene in 81–86° yield when catalyzed by chloro(triphenylphosphine)ruthenium(II). In the presence of cuprous chloride, dibromoacetonitrile also adds to isoprene (69). Isoprene reacts with carbon tetrabromide in carbon tetrachloride solvent on Co irradiation (70). Addition is induced using bromotrichloromethane in a 3:1 molar ratio to isoprene by x-irradiation (71). Only small amounts of the 1,2- and 4,3-addition products are obtained. The main reaction yields a mixture of ca 75° 1,4- and 25° 4,1-addition product.

Hydrocarbons. The reaction of isoprene with toluene, ethylbenzene, or isopropylbenzene is catalyzed by sodium or potassium (72). The reactions are carried out at 125°C in a pressure autoclave by adding the isoprene slowly to the alkylarene in which the alkali metal is dispersed along with a trace quantity of *o*-chlorotoluene which is used as a chain initiator. The products are chiefly monopentenylated in the side chain, and no information can be obtained on whether the addition is 1,4- or 1,2- since under these conditions the double bond migrates. The alkene products subsequently are reduced to alkanes by hydrogenation using 5° palladium on charcoal as catalyst.

With a sodium catalyst, the head product-to-tail product ratios are 2.77:1, 1.88:1, and 1.98:1 for toluene, ethylbenzene, and isopropylbenzene, respectively; corresponding ratios obtained using potassium are 3.04:1, 2.54:1, and 3.30:1.



head product



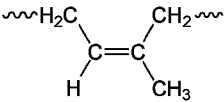
tail product

Alkylation of cyclohexane with isoprene can be carried out with alkyl radicals formed at 450°C and 20.3 MPa (200 atm) (73). 40° Pentenylcyclohexanes, 20° dipentenenes (ie, substances having the general formula C₁₀H₁), and 40° higher boiling compounds are obtained using a 6.8 molar ratio of cyclohexane to isoprene and a space velocity of 2.5 h⁻¹. Of the pentenylcyclohexanes, the head and tail products are in equal amounts. Even stable radicals, eg, triphenylmethyl, add readily to isoprene (74). Olefins, eg, ethylene, propylene, and styrene, add to isoprene in the presence of coordination catalysts that are based on cobalt, nickel, or iron (8).

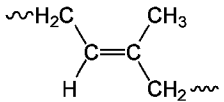
Other Compounds. Primary and secondary amines add 1,4- to isoprene (75). For example, dimethylamine in benzene reacts with isoprene in the presence of sodium or potassium to form dimethyl(3-methyl-2-butenyl)amine. Similar results are obtained with diethylamine, pyrrolidine, and piperidine. Under the same conditions, aniline and *N*-methylaniline do not react. Isoprene reacts with phenol in the presence of aluminum phenoxide (76) or concentrated phosphoric acid (77) to give complex products.

At 165°C and in the presence of chloroplatinic acid as catalyst, isoprene reacts with trichlorosilane, methylchlorosilane, ethylchlorosilane, benzylchlorosilane, and dibenzylchlorosilane (72). The addition is 1,4- with the substituted silane group attaching to the first carbon atom. Trimethylsilane does not react under these conditions. However, under similar conditions, heptamethylcyclotetrasiloxane reacts with isoprene by 1,2-addition (78). Thiophene reacts with isoprene in the presence of phosphoric acid to give mainly 2-(3-methyl-2-butenyl)thiophene and some higher boiling compounds (79). Reactions of isoprene with Grignard reagents have been described (80,81).

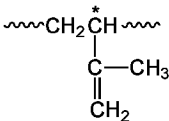
Polymerization. Isoprene polymerization can proceed by either 1,4- or 1,2-(vinyl)addition (see ELASTOMERS, SYNTHETIC POLYISOPRENE). 1,4-Addition leads to two possible structures which differ in the configuration of the remaining double bond. Vinyl addition produces two other possible structures, depending on whether the 1,2- or the 3,4- double bond takes part in the polymerization reaction.



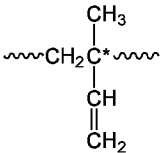
cis-1,4



trans-1,4



3,4-



1,2-

Any of the four monomer residues can be arranged in a polymer chain in either head-to-head, head-to-tail, or tail-to-tail configurations. Each of the two head-to-tail vinyl forms can exist as syndiotactic or isotactic structures because of the presence of an asymmetric carbon atom (marked with an asterisk) in the monomer unit. Of course, the random mix of syndiotactic and isotactic, ie, atactic structures also exists. Of these possible structures, only

three are known, *cis*-1,4-polyisoprene is exemplified by natural rubber (hevea, guayule) and synthetic analogues (see RUBBER, NATURAL). Balata and gutta-percha are natural *trans*-1,4-polyisoprene, and there is synthetic trans as well. A high 3,4-polyisoprene has been prepared (82) and is amorphous, thus it is considered atactic. A crystalline 3,4-polyisoprene was reported in 1993 by Goodyear (83). Chinese scientists have reported similar results and the nmr of this 3,4-polyisoprene. Stereochemistry was not assigned due to the complex nature of the nmr spectrum. Because the polymer is crystalline and has a T_{m} of 120–130°C, the Goodyear investigators have assigned it the syndiotactic structure. There are still no reports of isotactic structure. None of the three possible 1,2-polyisoprenes has been synthesized. Many polymers with mixed cis, trans, or vinyl structures have been prepared. Several reports of mixed head-to-head and head-to-tail chains have appeared (84).

The physical properties of any polyisoprene depend not only on the microstructural features but also on macro features such as molecular weight, crystallinity, linearity or branching of the polymer chains, and degree of cross-linking. For a polymer to be capable of crystallization, it must have long sequences where the structure is completely stereoregular. These stereoregular sequences must be linear structures composed exclusively of 1,4-, 1,2-, or 3,4-isoprene units. If the units are 1,4- then they must be either all cis or all trans. If 1,2- or 3,4- units are involved, they must be either syndiotactic or isotactic. In all cases, the monomer units must be linked in the head-to-tail manner (85).

Al–Ti Catalyst for cis-1,4-Polyisoprene. Of the many catalysts that polymerize isoprene, four have attained commercial importance. One is a coordination catalyst based on an aluminum alkyl and a vanadium salt which produces *trans*-1,4-polyisoprene. A second is a lithium alkyl which produces 90% *cis*-1,4-polyisoprene. Very high (99%) *cis*-1,4-polyisoprene is produced with coordination catalysts consisting of a combination of titanium tetrachloride, TiCl₄, plus a trialkylaluminum, R₃Al, or a combination of TiCl₄ with an alane (aluminum hydride derivative) (86–88).

Table 3 provides typical specifications for isoprene that are suitable for Al–Ti polymerization (89). Traditional purification techniques including superfractionation and extractive distillation are used to provide an isoprene that is practically free of catalyst poisons. Acetylenes and 1,3-cyclopentadiene are the most difficult to remove, and distillation can be supplemented with chemical removal or partial hydrogenation. Generally speaking distillation is the preferred approach. Purity is not the main consideration because high quality polymer can be produced from monomer with relatively high levels of olefins and *n*-pentane. On the other hand, there must be less than 1 ppm of 1,3-cyclopentadiene.

Table 3. Isoprene Specification for the Preparation of High *cis*-Polyisoprene^a

Component	Specification, ppm ^b
isoprene	99.0 wt % (min)
isoprene dimer	0.2 wt %
olefins	1.0 wt %
acetylenes	30
allenes	100
carbonyls	30
sulfur	25
piperylene	100
cyclopentadiene	1.0
peroxides	3
acetonitrile	10

^a Ref. 89.
^b Unless otherwise noted.

Alkali Metal Catalysts. The polymerization of isoprene by alkali metal and organometallic compounds (90–92) (other than organolithium) is a heterogeneous reaction both in bulk and hydrocarbon solvents. A homogeneous reaction takes place only in the presence of highly polar solvents. A comprehensive evaluation of the influence of solvent and positive counterion on polymer structure has been given (93). Only lithium-based initiators in hydrocarbon lead to high cis-1,4 structures. A trend of decreasing cis structures is observed with an increase in solvent basicity. Alkali metals, other than lithium, generally yield polymers of mixed structures with low cis content. The ionic character of the propagating ion pair depends mainly on the metal counterion and the solvent type.

The first successful use of lithium metal for the preparation of a *cis*-1,4-polyisoprene was announced in 1955 (50); however, lithium metal catalysis was quickly phased out in favor of hydrocarbon soluble organolithium compounds. These initiators provide a homogeneous system with predictable results. Organolithium initiators are used commercially in the production of *cis*-1,4-polyisoprene, isoprene block polymers, and several other polymers.

The polymerization of isoprene in hydrocarbon solvents with organo(mono)lithium compounds, eg, butyllithium, is a homogeneous reaction. In the absence of inhibiting impurities, the reaction starts immediately and proceeds to essentially 100% conversion. At moderate temperatures, in hydrocarbon solvent and in the absence of compounds having an active hydrogen, there is practically no chain termination or transfer. This polymerization lends itself to the preparation of linear polymers of controllable molecular weights and narrow molecular weight distribution. Monomer of very high quality is necessary, but the polymerization is more tolerant of 1,3-cyclopentadiene than Ti–Al. Soluble organolithium compounds are the only initiators used in the preparation of block copolymers containing isoprene or butadiene. The failure of Ziegler–Natta catalysts to produce block copolymer resulted in the use of organolithium catalyst for the polymerization of diblock, triblock, and tetrablock copolymers of styrene and butadiene. The preparation of SIS (styrene isoprene styrene) and SISIS block polymers spurred new uses for isoprene monomer in the adhesive market. The introduction of block copolymers using 70–80% styrene and 20–30% isoprene or butadiene opened new film applications for these polymers.

Other Polymerization Systems. Extensive work during the 1970s on polymerization of butadiene in solution with soluble nickel complexes has led to general acceptance of the idea that catalyst sites involve a monometallic π -complex with a monomer (94,95). A number of other coordination catalysts may be used to form high or very high *cis*-1,4-polyisoprene, eg, Zr salts or Mg alkyls (96), a cerium salt-based catalyst (97,98), and uranium- and thorium-based catalysts (99). Uranium tetraalkoxide–aluminum alkyl–Lewis acid combinations can be used to form high *cis*-1,4-polybutadiene and high *cis*-1,4-polyisoprene (100). Several of the lanthanide and actinide rare-earth series make active catalysts which produce very high cis-1,4 structures from both monomers.

Another group of isoprene polymerization catalysts is based on alanes and TiCl₄. In place of alkyl aluminum, derivatives of AlH₃ (alanes) are used and react with TiCl₄ to produce an active catalyst for the polymerization of isoprene. These systems are unique because no organometallic compound is involved in producing the active species from TiCl₄. The substituted alanes are generally complexed with donor molecules of the Lewis base type, and they are liquids or solids that are soluble in aromatic solvents. The performance of catalysts prepared from AlHCl₂O(C₂H₅)₂ with TiCl₄ has been reported (101).

The preparation of high *trans*-1,4-polyisoprene with VCl₃ plus (C₂H₅)₃Al catalyst has been described; it has been concluded that there are several species of active sites all of which give high *trans*-polyisoprene. Similar conclusions have been reached about varying catalyst species with vanadium salts during EPDM polymerizations. Several vanadium salts (VCl₃, VOCl₃, VCl₂) all yield *trans*-polyisoprene catalysts which might argue that the active catalyst is a single, lower valent species. *trans*-Polyisoprene also has been made with a η^3 -2-propyl iodonickel catalyst (with no co-catalyst) and with tris(η^3 -2-propenyl) chromium that is deposited on aluminosilicate (102).

Production

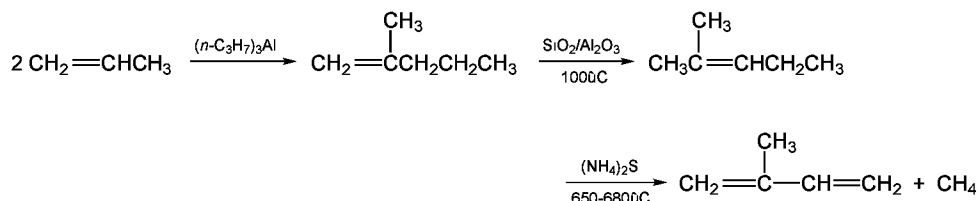
The usage of isoprene monomer is somewhat limited by price and availability. The historical large usage has been in the production of *cis*-1,4-polyisoprene

which is similar to natural rubber. Although the price of natural rubber was expected to increase over time, it has remained relatively inexpensive since the 1960s. As a result, polyisoprene has been produced primarily where nonprice considerations were of utmost importance. The largest capacity has been and remains in the CIS (former USSR) region. Several plants around the world have been shut down, and the trend appears to continue downward. On the other hand the use of isoprene in block copolymers has grown rapidly. This growth has tended to offset some of the decline of *cis*-1,4-polyisoprene. In 1992 isoprene monomer was produced in Brazil, the Netherlands, Japan, CIS, Romania, and the United States. *cis*-1,4-Polyisoprene was produced in all locations except Brazil. World capacity for *cis*-1,4-polyisoprene is difficult to gauge. It is probably about 1,150,000 t, but the operating rate is probably well below that level. This compares with production in 1976 of 650,000 t. Because most producers of isoprene also produce polymer, accurate figures are not readily available.

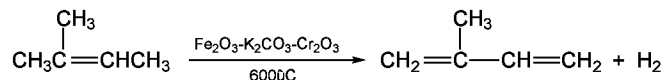
The principal route for production of isoprene monomer outside of the CIS is recovery from ethylene by-product C₅ streams. This route is most viable where ethylene is produced from naphtha or gas oil and where several ethylene plants are located in relatively close proximity to the isoprene plant. Although the yield of isoprene per mass of ethylene is quite low, there is enough ethylene produced to provide a large portion of demand. Because of the presence of *n*-pentane in these streams which azeotropes with isoprene, extractive distillation must be used to recover pure isoprene. Acetonitrile is the most common solvent, but dimethylformamide is also used commercially.

Synthesis. Because of the limited availability of by-product isoprene much effort has been devoted to synthesis of isoprene. Most routes tend to have marginal selectivity and require large amounts of energy. The choice of which route is preferable depends on availability and cost of raw materials and cost of energy. Several synthetic routes have been practiced commercially (103–108).

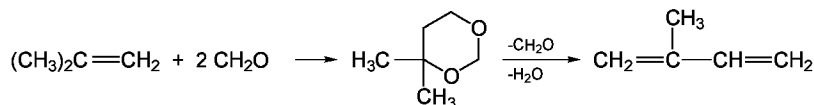
Propylene Dimer. The synthesis of isoprene from propylene (109,110) is a three-step process. The propylene is dimerized to 2-methyl-1-pentene, which is then isomerized to 2-methyl-2-pentene in the vapor phase over silica alumina catalyst. The last step is the pyrolysis of 2-methyl-2-pentene in a cracking furnace in the presence of (NH₄)₂S (111,112). Isoprene is recovered from the resulting mixture by conventional distillation.



Dehydrogenation of Tertiary Amylenes. The starting material here is a C₅ fraction which is cut from catalytic cracking of petroleum. Two of the tertiary amylene isomers, 2-methyl-1-butene and 2-methyl-2-butene, are recovered in high purity by formation of methyl tertiary butyl ether and cracking of this to produce primarily 2-methyl-2-butene. The amylenes are mixed with steam and dehydrogenated over a catalyst. The crude isoprene can be purified by conventional or extractive distillation.

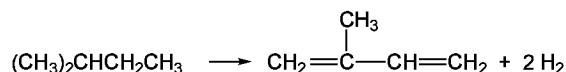


Isobutylene–Formaldehyde. Isobutylene is condensed with formaldehyde at 95°C to give the principal product 4,4-dimethyl-*m*-dioxane. In the second step, the dioxane is decomposed in the presence of an acid catalyst to isoprene, formaldehyde, and water.

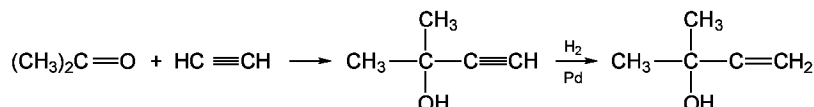


Much of the work with regard to this process was done by the French Petroleum Institute (113) and by the Kuraray Co. (108). In the CIS, a similar process which begins with crude C₅s was developed (114). A one-step process that begins with isobutylene and methanol has been disclosed (108,115). This process is believed to have significant economic advantages over the original route.

Isopentane Dehydrogenation. In isopentane dehydrogenation, which is used in the CIS, isopentane or a C₅ fraction from a catalytic cracker is dehydrogenated to isoprene (6):



Acetone–Acetylene. A process based on acetone and acetylene (113–118) first was utilized in Germany. ANIC of Italy and Karbochem of South Africa have used this process to produce isoprene:



The ethynylation reaction takes place at 10–40°C and 2 MPa (20 atm) and liquid ammonia is the solvent. The methylbutynol is converted into methylbutenol by selective hydrogenation and then is dehydrated over alumina at 250–300°C. Polymerization-grade isoprene is obtained.

Butene Hydroformylation. A more recently developed process for the synthesis of isoprene is butene hydroformylation followed by dehydration. This process has not been practiced commercially, but processing steps are similar to commercial processes (119). 2-Butene is hydroformylated to 2-methylbutanol which is then dehydrated to isoprene.

Health and Safety

Isoprene is not known to present serious toxicological hazards in handling (2); however, like many chemicals studies are ongoing. A concentration of 2° isoprene in air does not narcotize mice but produces bronchial irritation. However, concentrations of 5° are fatal to mice. In humans, a one minute inhalation of 0.16 mg isoprene per liter air is mildly irritating to the mucous membranes of the eyes, nose, and upper respiratory passages (120). It was proposed that the limit of isoprene concentration on industrial sites be set at 0.04 mg L air; it was also recommended that the maximum concentration of isoprene in water be set at 0.005 mg L. The extent of toxic conditions and air pollution by isoprene in the manufacturing areas of synthetic rubber and their vicinity has been dealt with in several CIS publications (121–123). LC₅₀ rat inhalation is 180 g m⁻² 2 h for rats, and prolonged exposure may cause central nervous system depression.

Isoprene is classified by the ICC as a flammable liquid requiring a red label (124). Its flash point is –54°C with a lower explosive limit (LEL) of 1.5°

and an upper explosive limit (UEL) of 8.9° o. It forms dangerous peroxides on exposure to air in the absence of inhibitors. A solution having 17 wt ° o of peroxide does not detonate in a standard drop test. The polymeric peroxide residue that is obtained on evaporation does explode. The reaction of isoprene and oxygen is rapid, with 1° o conversion of isoprene in about 3 h at 50°C; the product (125) is an alternating copolymer of oxygen and isoprene, with the repeating structure being $-(C_5H_5O_2)-$. The peroxide could serve as a source of initiation, which could lead to uncontrolled polymerization either homogeneously or as a popcorn growth, ie, peroxide polymer having the appearance of popcorn.

Because of the potential hazards on its exposure to oxygen, isoprene should be stored in an inert atmosphere (nitrogen) in the presence of at least 50 ppm of *t*-butylcatechol. Because the inhibitor is slowly consumed during storage, it is advisable to analyze the isoprene periodically and to add more inhibitor as needed. Before use it should be flash distilled to remove dimer and inhibitor. In industrial use, inhibitor is often removed by a caustic wash. A dangerous reaction of isoprene with ozone has been reported (126): when one gram of isoprene that was diluted with 50 mL of *n*-heptane was treated with ozone at -78°C, the resulting product exploded shortly after being removed from the cooling bath; however, the product of a similar reaction that was carried out at room temperature did not explode. On storage, isoprene forms cyclic dimers at a slow rate which is not affected by the presence of an inhibitor (35).

Economic Aspects and Applications

Isoprene pricing tends to vary considerably due to a fairly thin commercial market. Because isoprene raw materials are primarily petroleum based and synthesis or recovery is energy intensive, most pricing is indexed to petroleum and energy. For large-scale applications monomer production is in tandem with application production. Generally isoprene availability is less than butadiene, and the price is higher. Isoprene is used where the unique properties of the products can command a premium over butadiene. Almost all isoprene produced is used for the preparation of polymers or copolymers. *cis*-Polyisoprene is the largest application with SIS block polymers being a rapidly growing secondary application. Butyl rubber is a significant third application. Table 4 provides isoprene demand by use for 1985 and 1992, and provides synthetic polyisoprene consumption by region for the same years. Natural rubber consumption was 4.8–5.5 million tons per year over the same time period.

Table 4. World Isoprene Demand and Polyisoprene Consumption, 10³ t

	1985	1992
<i>Monomer use</i>		
polyisoprene	1170	827
SIS	36	95
butyl rubber	31	25
other	4	10
<i>Polymer use</i>		
North America	45	60
Latin America	5	2
Western Europe	40	40
Middle East and Africa	20	0
Far East and Australia	60	45
Eastern Europe, CIS, and China	1000	680

BIBLIOGRAPHY

"Isoprene" in *ECT* 2nd ed., Vol. 12, pp. 64–83, by G. Holden and R. H. Mann, Shell Chemical Co.; in *ECT* 3rd ed., Vol. 13, pp. 818–837, by W. M. Saltman, The Goodyear Tire and Rubber Co.

1. C. G. Williams, *Proc. Roy. Soc.* **10**, 516 (1860); *Phil. Trans.*, 241 (1860).
2. G. Bouchardat, *Comp. Rend.* **80**, 1117 (1879).
3. U.S. Pat. **3,144,743** (Dec. 17, 1963), S. E. Horne, (to Goodrich-Gulf Co.).
4. S. E. Horne and co-workers, *Ind. Eng. Chem.* **48**, 784 (1956).
5. F. W. Stavely and co-workers, *Ind. Eng. Chem.* **48**, 778 (1956).
6. I. V. Garmonov, paper No. 1, presented at the *International Symposium on Isoprene Rubber*, Moscow, Russia, 1972.
7. W. A. Tilden, *Chem. News* **46**, 120 (1882); W. A. Tilden, *J. Chem. Soc.* **45**, 410 (1884).
8. W. J. Bailey, in E. C. Leonard, ed., *Vinyl and Diene Monomers*, Part II, John Wiley & Sons, Inc., New York, 1971, Chapt. 5.
9. R. W. Gallant, *Hydrocarbon Process. Pet. Ref.* **46**(a), 155 (1967); *Physical Properties of Hydrocarbons*, Vol. 1, Gulf Publishing Co., Houston, Tex., 1968, pp. 157–166.
10. R. J. W. Le Feure and K. M. S. Sundaran, *J. Chem. Soc.*, 3547 (1963); R. J. W. Le Feure and K. M. S. Sundaran, *J. Chem. Soc.*, 3518 (1964).
11. L. V. Vilkov and I. N. Sadova, *Zh. Strukt. Khim.* **8**, 398 (1967).
12. G. J. Szasz and N. Sheppard, *Trans. Faraday Soc.* **49**, 358 (1953); M. I. Batuev, A. S. Onischenko, A. D. Matveeva, and N. I. Aronova, *Dokl. Akad. Nauk. SSSR* **132**, 581 (1960); D. Craig, J. J. Shipman, and R. B. Fowler, *J. Am. Chem. Soc.* **83**, 2885 (1961); H. Dodziuk, *J. Mol. Struct.* **20**, 317 (1974).
13. J. Gresser, A. Rajbenbach, and M. Szwarc, *J. Am. Chem. Soc.* **82**, 5820 (1960).
14. D. A. C. Compton, W. O. George, and W. F. Maddams, *J. Chem. Soc. (Perkins II)* **14**, 1666 (1976).
15. S. Dzhezzati, A. R. Kyazimova, V. I. Tyulin, and Yu. A. Pentin, *Vestn. Mosk. Univ. Khim.* **23**(5), 19 (1968).
16. L. Pauling, *The Nature of the Chemical Bond*, 3rd ed., Cornell University Press, Ithaca, N.Y., 1960, p. 291.
17. R. Hoffmann and R. A. Olofson, *J. Am. Chem. Soc.* **88**, 943 (1966).
18. R. S. Stearns and L. E. Forman, *J. Polym. Sci.* **41**, 381 (1959).
19. I. Kuntz and A. Gerber, *J. Polym. Sci.* **42**, 299 (1960).
20. R. J. Orr, *J. Polym. Sci.* **58**, 843 (1962).
21. M. Szwarc, *J. Polym. Sci.* **40**, 583 (1959).
22. R. Adams, ed., *Organic Reactions*, Vol. 4, John Wiley & Sons, Inc., New York, 1948, pp. 60–173.
23. E. F. Lutz and G. M. Bailey, *J. Am. Chem. Soc.* **86**, 3899 (1964).
24. Yu. A. Titov and A. I. Kuznetsova, *Izv. Akad. Nauk SSSR Otd. Khim. Nauk*, 1297 (1960).
25. I. N. Nazarov, Yu. A. Titov, and A. R. Kuznetsova, *Izv. Akad. Nauk SSSR Otd. Khim Nauk*, 1412 (1959).
26. H. E. Hennis, *J. Org. Chem.* **28**, 2570 (1963).
27. A. N. Volkov, A. V. Bogdanova, and M. F. Shostakovskii, *Izv. Akad. Nauk SSSR Otd. Khim. Nauk*, 1280 (1962).
28. M. F. Shostakovskii, A. V. Bogdanova, and A. N. Volkov, *Izv. Akad. Nauk SSSR Otd. Khim Nauk*, 1284 (1962).
29. F. Bergmann and H. E. Eschinazi, *J. Am. Chem. Soc.* **65**, 1405 (1943).
30. D. Craig, J. J. Shipman, and R. B. Fowler, *J. Am. Chem. Soc.* **83**, 2885 (1961).

31. E. Gil-AV and Y. Herzberg-Minzly, *Proc. Chem. Soc.*, 316 (1961).
32. A. F. Plate and N. A. Belikova, *Zh. Obshch. Khim.* **30**, 3953 (1960).
33. A. S. Onishchenko, *Diene Synthesis*, Daniel Davey and Co., Inc., New York, 1964, pp. 592–595.
34. Ya. M. Paushkin, A. G. Liakumovich, Yu. I. Michurov, R. B. Valitov and A. F. Lunin, *Tr. Mosk. Inst. Neftekhim. Gazov. Prom.* (72), 23 (1967).
35. *Storage and Handling of Liquefied Olefins and Diolefins*, Enjay Chemical Co., New York, 1962, p. 19.
36. G. S. Hammond, N. J. Turro, and R. S. H. Liu, *J. Org. Chem.* **28**, 3297 (1963).
37. R. S. H. Liu, N. J. Turro, and G. S. Hammond, *J. Am. Chem. Soc.* **87**, 3406 (1965).
38. B. S. Greensfelder and H. H. Voge, *Ind. Eng. Chem.* **37**, 983 (1945).
39. L. I. Zatharkin, *Dokl. Akad. Nauk SSSR* **131**, 1069 (1960).
40. H. Takahashi and M. Yagamuchi, *Osaka Kogyo Gijutsu Shikensho Kiho* **15**, 271 (1964).
41. L. I. Zakharkin and G. G. Zhigareva, *Izv. Akad. Nauk SSSR Ser. Khim.*, 168 (1964).
42. J. P. Candlin and W. H. Janes, *J. Chem. Soc. C*, 1856 (1968).
43. A. Misono, Y. Uchida, M. Hidai, and Y. Ohsawa, *Bull. Chem. Soc. Jpn.* **39**, 3425 (1966).
44. R. F. Heck, *Organotransition Metal Chemistry*, Academic Press, Inc., New York, 1974, p. 162.
45. R. L. Frank and R. P. Seven, in E. C. Horning, ed., *Organic Syntheses*, Collective Vol. III, John Wiley & Sons, Inc., New York, 1955, p. 499.
46. R. L. Frank, C. E. Adams, J. B. Blegen, R. Deanin, and P. V. Smith, *Ind. Eng. Chem.* **39**, 887 (1947).
47. T. Akiyama, M. Sugihara, T. Imagawa, and M. Kawanisi, *Bull. Chem. Soc. Jpn.* **51**, 1251 (1978).
48. H. Takahashi and M. Yamaguchi, *Osaka Kogyo Gijutsu Shikensho Kiho* **15**, 271 (1964).
49. **Fr. Pat. 1,393,071** (Mar. 19, 1965), A. Carbonaro, A. Bonfardeci, and L. Porri (to Montecatini).
50. **U.S. Pat. 3,429,940** (Feb. 25, 1969), F. T. Wadsworth (to Columbian Carbon Co.).
51. J. M. Pearson and M. Szwarc, *Trans. Faraday Soc.* **60**, 553 (1964).
52. A. Rajbenbach and M. Szwarc, *Proc. Roy. Soc. (London)* **A251**, 394 (1959).
53. D. J. Trecker, R. L. Brandon, and J. P. Henry, *Chem. Ind. (London)*, 652 (1963).
54. K. J. Crowley, *Proc. Chem. Soc.* **334** (1962); *Tetrahedron* **21**, 1001 (1965).
55. R. Srinivasan, *J. Am. Chem. Soc.* **84**, 4141 (1962).
56. A. A. Oswald, K. Griesbaum, W. Thaler, and B. E. Hudson, Jr., *Am. Chem. Soc. Div. Petrol. Chem. Prepr.* **7**, 139 (1962).
57. M. J. Ljam, *The Addition of Halogens to Chloroprene and Isoprene*, Ph.D. dissertation, The University of Texas, Austin, 1961, p. 110.
58. E. G. E. Hawkins and M. D. Philpot, *J. Chem. Soc.*, 3204 (1962).
59. A. A. Petrov, *J. Gen. Chem. USSR* **13**, 741 (1943).
60. A. J. Ultee, *J. Chem. Soc.*, 530 (1948).
61. W. J. Jones and H. W. T. Cleorley, *J. Chem. Soc.*, 832 (1946).
62. V. N. Belov, V. K. Promonenkov, and A. B. Kamenskii, *Zh. Obshch. Khim.* **34**, 3432 (1964).
63. A. Ledwith and R. M. Bell, *Chem. Ind. (London)*, 459 (1959).
64. E. C. Herrick, *J. Org. Chem.* **24**, 139 (1959).
65. P. D. Bartlett, *J. Am. Chem. Soc.* **86**, 616 (1964).
66. Y. Kimura, I. Isagawa, and Y. Otsuji, *Chem. Lett.* (8), 951 (1977).
67. A. A. Petrov, N. A. Razumova, and M. L. Genusov, *Zh. Obshch. Khim.* **28**, 1128 (1958).
68. H. Matsumoto and co-workers, *Chem. Lett.* (1), 115 (1978).
69. **Fr. Pat. 1,409,516** (Aug. 27, 1965), H. W. Moore, F. F. Rust, and H. S. Klein (to Shell).
70. G. Rabilloud and P. Traynard, *Compt. Rend.* **251**, 1505 (1960).
71. C. S. H. Chen and E. F. Hosterman, *J. Org. Chem.* **28**, 1585 (1963).
72. L. Shiihara, W. F. Hoskyns, and H. W. Post, *J. Org. Chem.* **26**, 4000 (1961).
73. L. Kh. Friedlin, N. M. Nazariyam, E. F. Litvin, and G. K. Gaivoronskaya, *Neftehimaya* **4**, 246 (1964).
74. J. B. Conant and H. W. Scherp, *J. Am. Chem. Soc.* **53**, 1941 (1931).
75. G. T. Martirosyan and E. A. Grigoryan, *Izv. Akad. Nauk Arm. SSR Khim. Nauk* **16**, 31 (1963).
76. K. C. Dewhirst and F. F. Rust, *J. Org. Chem.* **28**, 798 (1963).
77. A. R. Bader and W. C. Bean, *J. Am. Chem. Soc.* **80**, 3073 (1958).
78. K. A. Andrianov and co-workers, *Dokl. Akad. Nauk SSSR* **158**, 133 (1964).
79. H. Pines, B. Kvetinskas, J. A. Vesely, and E. Baclawski, *J. Am. Chem. Soc.* **73**, 5173 (1951).
80. M. S. Kharasch, R. D. Mulley, and W. Nudenberg, *J. Org. Chem.* **19**, 1477 (1954).
81. M. S. Kharasch, P. G. Holten, and W. Nudenberg, *J. Org. Chem.* **19**, 1600 (1954).
82. G. Natta, L. Porri, and A. Carbonaro, *Makromol. Chem.* **77**, 126 (1964).
83. W. L. Hsu and A. F. Halasa, *144th Meeting of the Rubber Division*, Paper No. 77, American Chemical Society, Washington, D.C., Oct. 1993.
84. E. Schoenberg, H. A. Marsh, S. J. Walters, and W. M. Saltman, *Rubber Chem. Tech.* **52**, 526 (1979).
85. G. Natta, *Mod. Plast.* **34**, 169 (1956).
86. C. T. Winchester, *Ind. Eng. Chem.* **51**, 19 (1959).
87. W. M. Saltman and E. Schoenberg, in J. R. Elliot, ed., *Macromolecular Syntheses*, Vol. 2, John Wiley & Sons, Inc., New York, 1966, p. 50.
88. **U.S. Pat. 2,851,505** (Sept. 9, 1958), S. M. Henke and V. N. Hurd (to Gulf Research and Development Co.); **Brit. Pat. 837,908** (1962), (to Shell); **U.S. Pat. 2,900,430** (Aug. 18, 1959), A. M. Henke and H. C. Stauffer (to Gulf Research and Development Co.).
89. R. Fowler and D. Barker, *Chem. Eng. (London)*, 322 (1971).
90. **Brit. Pat. 24,790** (Oct. 25, 1911), F. E. Matthews and E. H. Strange.
91. C. H. Harries, *Ann.* **383**, 184 (1911).
92. **Ger. Pats. 255,786** (Jan. 27, 1912), and 255,787 (Apr. 9, 1912), I. H. Labhardt (to BASF Badische Anilin).
93. R. V. Tobolsky and C. E. Rogers, *J. Polym. Sci.* **40**, 73 (1959).
94. Ph. Teyssie and F. Dawans, in W. M. Saltman, ed., *The Stereo Rubbers*, Wiley-Interscience, New York, 1977, Chapt. 3.
95. B. A. Dolgoplosk, *Kinet. Katal.* **18**, 1146 (1977).
96. W. Cooper, in Ref. 94, pp. 50–53.
97. **U.S. Pat. 3,297,667** (Jan. 10, 1967), W. C. VonDohlen, T. P. Wilson, and E. G. Cafilisch (to Union Carbide).
98. **U.S. Pat. 3,541,063** (Nov. 17, 1970), M. C. Throckmorton and W. M. Saltman (to The Goodyear Tire and Rubber Co.).
99. **U.S. Pat. 3,676,411** (July 11, 1972), M. C. Throckmorton and W. M. Saltman (to The Goodyear Tire and Rubber Co.).
100. M. Burzzone, A. Mazzei, and G. Giuliani, *Rubber Chem. Technol.* **47**, 1175 (1974).
101. W. Marconi, A. Mazzei, S. Cesca, and M. deMalde, *Chim. i Ind. (Milan)* **51**, 1084 (1969); W. Marconi, A. Mazzei, S. Cucinella, and M. de Malde, *Makromol. Chem.* **71**, 118, 134 (1964).
102. I. M. Kosheleva, N. N. Stefanovskaya, and V. L. Shmonena, *Kautsch. i Rezina*. **6**, 49 (1975).
103. M. J. Rhoad, *Proceedings of the 15th Annual Meeting*, International Institute of Synthetic Rubber Producers, Inc., Kyoto, Japan, 1974; *Rubber Ind. (London)* **9**(12), 68 (1975); *Rubber Plast. News*, 14 (Jan. 23, 1978).
104. S. K. Ogorohnikov and G. S. Idlis, *Isoprene Production*, Izdatelstvo Khimia, Leningrad, Russia, 1973.
105. *Isoprene*, Report No. 28, Stanford Research Institute, Menlo Park, Calif., 1967.

106. C. Capitani, *Rev. Gen. Caout. Plast.* **51**(4), 195 (1974).
107. A. Mitsutani, *Yuki Gosei Kagaku Kyokai Shi* **32**, 528 (1974).
108. A. Mitsutani and S. Kumano, *Chem. Econ. Eng. Rev.* **3**(2), 35 (1971).
109. V. J. Anhorn, K. J. Frech, J. J. Tazuma, P. H. Wise, and W. E. Morrow, *Chem. Eng. Prog.* **57**(5), 41 (1961).
110. V. J. Anhorn, K. J. Frech, G. S. Schaffel, and D. Brown, *Rubber Plast. Age* **42**, 1212 (1961).
111. H. J. Osterhof, *Rev. Gen. Caout. Plast.* **42**, 529 (1965).
112. **U.S. Pat. 3,284,532** (Nov. 8, 1966), K. J. Frech (to The Goodyear Tire and Rubber Co.).
113. *Hydrocarbon Process. Pet. Ref.* **92**(11), 187 (1963); *Hydrocarbon Process.* **50**(11), 167 (1971).
114. *Hydrocarbon Process.* **50**(11), 168 (1971).
115. H. J. Peterson and J. O. Tucker, *Hydrocarbon Process.* **53**(7), 21 (1974).
116. *Hydrocarbon Process.* **50**(11), 170 (1971).
117. *Rubber World* **167**(2), 16 (1972).
118. M. DeMalde, *Chim. e Ind. (Milan)* **45**, 665 (1963).
119. **U.S. Pat. 4,524,233** (June 18, 1985), W. L. Hsu and co-workers (to The Goodyear Tire and Rubber Co.).
120. V. D. Gostinskii, *Gigiena Truda i Prof. Zabolevaniya* **9**(1), 36 (1965).
121. V. I. D'yachkov, V. K. Mishin, L. N. Berezina, and P. A. Podurueva, *Sb. Nauch. Tr. Kuibyshev. Nauch. Issled. Inst. Gig.* (7), 94 (1972).
122. S. A. Pigolev, *Sb. Nauch. Tr. Kuibyshev. Nauch. Issled. Inst. Gig.* (6), 64 (1971).
123. A. S. Fauston, *Tr. Voronezh. Med. Ist.* (87), 10 (1972).
124. *Isoprene Technical Data*, Bulletin No. 1, Goodrich-Gulf Chemicals, Inc., Cleveland, Ohio, 1968.
125. D. G. Hendry, Stanford Research Institute, Menlo Park, Calif., private communication, 1972.
126. *Chem. Eng. News* **34**, 292 (1956).

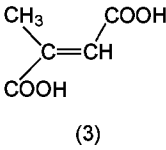
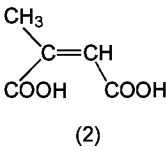
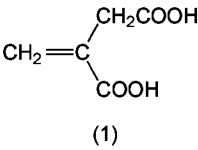
Hugh M. Lybarger
The Goodyear Tire and Rubber Company

ITACONIC ACID AND DERIVATIVES

Itaconic acid[97-65-4] (methylenebutanedioic acid, methylenesuccinic acid) is a crystalline, high melting acid (mp 167–168) produced commercially by fermentation of carbohydrates (1–4). Itaconic acid is produced in the broth from citric acid (qv). Isolated from the pyrolysis products of citric acid in 1836, this α -substituted acrylic acid received its name by rearrangement of aconitic, the acid from which it is formed by decarboxylation.

Physical and Chemical Properties

The dissociation constants in H₂O at 25°C of itaconic acid are as follows: $K_1 = 1.40 \times 10^{-4}$; $K_2 = 3.56 \times 10^{-6}$ (5). Heat of formation is 840 kJ/mol (201 kcal/mol) and heat of combustion is 1980 kJ/mol (473 kcal/mol) (6). Its solubility is 9.5 g/100 mL H₂O at 25°C and increases with temperature and pH (7). Itaconic acid (1) is isomeric with citraconic [498-23-7] (2) and mesaconic [498-24-8] (3) acids. Under acidic, neutral, or mildly basic conditions and at moderate temperatures, itaconic acid is stable. At elevated temperatures or under strongly basic conditions, the isomers are interconvertible.



Itaconic acid, anhydride, and mono- and diesters undergo vinyl polymerization. Rates of polymerization and intrinsic viscosities of the resulting homopolymers are lower than those of the related acrylates (see ACRYLIC ESTER POLYMERS) (8,9).

Uses

Itaconic acid is a specialty monomer that affords performance advantages to certain polymeric coatings (qv) (see POLYESTERS, UNSATURATED). Emulsion stability, flow properties of the formulated coating, and adhesion to substrates are improved by the acid. Acrylonitrile fibers with low levels of the acid comonomer exhibit improved dye receptivity which allows more efficient dyeing to deeper shades (see ACRYLONITRILE POLYMERS; FIBERS, ACRYLIC) (10,11). Itaconic acid has also been incorporated in PAN precursors of carbon and graphite fibers (qv) and into ethylene ionomers (qv) (12).

BIBLIOGRAPHY

"Itaconic Acid" in *ECT* 1st ed., Vol. 8, pp. 105–107, by C. J. Knuth, Chas. Pfizer & Co., Inc.; in *ECT* 2nd ed., Vol. 12, pp. 83–86, by C. J. Knuth, Chas. Pfizer & Co., Inc.; in *ECT* 3rd ed., pp. 865–873, by B. E. Tate, Pfizer Inc.

1. **U.S. Pat. 2,385,283** (Sept. 18, 1945), J. H. Kane, A. C. Finlay, and P. F. Amann (to Chas. Pfizer & Co., Inc.).
2. V. F. Pfeifer, C. Vojnovich, and E. N. Heger, *Ind. Eng. Chem.* **44**, 2975 (1952).
3. **U.S. Pat. 3,165,582** (Dec. 22, 1964), M. Batti (to Miles Laboratories, Inc.).
4. **U.S. Pat. 3,044,941** (July 17, 1962), R. C. Nubel and E. J. Ratajak (to Chas. Pfizer & Co., Inc.).
5. H. W. Ashton and J. R. Partington, *Trans. Faraday Soc.* **30**, 598 (1934).
6. R. C. Wilhoit and I. Lei, *J. Chem. Eng. Data* **10**(2), 166 (1965).
7. B. E. Tate, in E. C. Leonard, ed., *Vinyl and Diene Monomers*, Wiley-Interscience, New York, 1970, Pt. 1, p. 212.
8. B. E. Tate, *Adv. Polymer Sci.* **5**, 215 (1967).
9. J. Velickovic and S. Vasovic, *Makromol. Chem.* **153**, 207 (1972).
10. A. A. Kharkharov and G. S. Saribekov, *Tekstil. Prom.* **26**(5), 74 (1966).
11. **Ital. Pat. 572,367** (Jan. 25, 1958), E. Condorelli and co-workers (to Laboratori Italiani di Ricerca Chimica (LIRC)).
12. R. W. Rees and D. J. Vaughan, *Polym. Prepr. Am. Chem. Soc. Div. Polym. Chem.* **6**, 296 (1965).

J ACID.

See DYES AND DYE INTERMEDIATES.

JASMIN, JASMINE.

See OILS, ESSENTIAL; PERFUMES.

JUNIPER.

See OILS, ESSENTIAL.

JUTE.

See FIBERS, VEGETABLE.

KENAF.

See CHEMURGY; FUELS FROM BIOMASS.

KETENES, KETENE DIMERS, AND RELATED SUBSTANCES

Ketenes are oxo compounds with cumulated carbonyl and carbon–carbon double bonds of the general structure R₁R₂C=C=O, where R₁ and R₂ may be any combination of hydrogen, alkyl, aryl, acyl, halogen, and many other functional groups. Ketenes with R₁ = _HR₂ ≠ H are sometimes called aldoketenes, those with R₁,R₂ ≠ H, ketoketenes. The S– and N– analogues of ketenes are called thioketenes (R₁R₂C=C=S) and ketenimines (R₁R₂C=C=NR), respectively.

The parent substance, ketene itself [463-51-4] is the only ketene to be manufactured in very large industrial quantities. Its principal applications are for the manufacture of acetic anhydride [108-24-7] and diketene [674-82-8]. The latter is an important organic intermediate used as the source of acetoacetic esters, amides, and anilides, which are widely used in the preparation of fine chemicals, pigments, drugs, and agrochemicals. Dimeric long-chain alkylketenes (C₁₂–C₂₀) are used in industrial quantities as paper sizing agents.

The chemistry of ketenes is dominated by their high reactivity: most of them are not stable under normal conditions, many exist only as transient species. Nucleophilic attack at the *sp*-carbon, [2 + 2] cycloadditions, and ketene insertion into single bonds are the most important and widely used reactions of such compounds.

Ketenes and related compounds have been reviewed extensively (1–9). For the synthesis and synthetic uses of conjugated ketenes see Reference 10. Ketenes with three or more cumulated double bonds have been prepared (11,12). The best known is carbon suboxide [504-64-3], C₃O₂, which has preparative uses and has been reviewed (13–16). Thioketenes (17,18), ketenimines (19–21), and their dimers show interesting reactivity, but they have not achieved industrial importance to date.

Monomeric Ketenes

Physical Properties. Ketenes range in their properties from colorless gases such as ketene and methylketene [6004-44-0] to deep colored liquids such as diphenylketene [525-06-4] and carbon subsulfide [627-34-9]. Table 1 lists the physical state mp, and bp for certain ketenes, thioketenes, and ketenimines.

Other important physical properties of the parent compound ketene are as follows. Density is 0.65 g mL at –60°C, whereas vapor density, compared to theoretical, air = 1, is 1.45. Free energy of formation Δ*G*_{*f*}[°] = –49.6 ± 1.6 kJ mol^{–1} and enthalpy of formation Δ*H*_{*f*}[°] = –47.7 kJ mol^{–1} (–11.4 kcal mol^{–1}). The dipole moment is 4.7 × 10^{–30} Cm (1.41 D).

Vapor pressure data from –71 to 90°C has been given:

vapor pressure, <i>t</i> , °C	–71°C	–20°C	–30°C	–40°C	–50°C	–90°C
vapor pressure, MPa	0.02	2.0	2.64	3.41	4.35	10.03

To convert MPa to mm Hg, multiply by 7500.

Chemical Properties. The chemistry of ketenes is dominated by the strongly electrophilic *sp*-hybridized carbon atom and a low energy lowest unoccupied molecular orbital (LUMO). Therefore, ketenes are especially prone to nucleophilic attack at C1 and to [2 + 2] cycloadditions. Less frequent reactions are the so-called ketene insertion, a special case of addition to substances with strongly polarized or polarizable single bonds (37), and the addition of electrophiles at C2. For a review of addition reactions of ketenes see Reference 8.

Nucleophilic Addition. Reagents with labile hydrogen atoms, such as alcohols, thiols, phenols, carboxylic acids and amines, add to ketenes giving the corresponding carboxylic acid derivatives (1) as shown in Figure 1 (38). Not many are of practical importance, as there are better ways to such

compounds. A notable exception is the manufacture of acetic anhydride from ketene and acetic acid [64-19-7], which is still of importance even though a new industrial process based on carbonylation of methyl acetate [79-20-9] appeared in the 1980s (39).

2 + 2 **Cycloaddition.** Ketenes are ideal components in [2 + 2] cycloadditions for additions to the opposite sides of a π -system as shown in the cyclobutane product (2) in Figure 1. Electron-rich double bonds react readily with ketenes, even at room temperature and without catalysts. In conjugated systems, ketenes add in a [2 + 2] fashion. This is illustrated in the reaction following, where the preferential orientation of L (large substituent) and S (small substituent) is seen (40). This reaction has been used in the synthesis of tropolone [533-75-5].

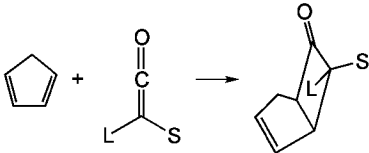


Table 1. Properties of Some Ketenes, Thioketenes, and Ketenimines

Name	CAS Registry Number	Physical state	Mp or bp, °C ^a
ketene	[463-51-4]	colorless gas	mp 134.1, bp 41 (22)
diphenylketene	[525-06-4]	orange liquid	mp 9–10, bp 96–99 (2.67 Pa) (23)
carbon suboxide	[504-64-3]	colorless gas	mp 112.2 (24), bp 6.8 (25)
dichloroketene	[4591-28-0]	not isolated	
dimethylketene	[598-26-5]	yellow liquid	mp 97.5, bp 34 (26)
bis(trifluoromethyl)ketene	[684-22-0]	colorless gas	bp 5 (27)
methylketene	[6004-44-0]	colorless gas	23 (28)
tert-butylcyanoketene	[29342-22-1]	isolated in solution only (29)	
(trimethylsilyl)ketene	[4071-85-6]	colorless oily liquid	bp 81–82 (30)
methylphenylketene	[3156-07-8]	liquid	mp 7, bp 76 (1.86 kPa) (31)
carbon subsulfide	[627-34-9]	dark red liquid	bp 60–40 (1.6 kPa)
thioketene	[18282-77-4]	detected, not isolated (32)	
bis(tert-butyl)thioketene	[16797-75-4]	purple liquid	bp 64–65 (0.8 kPa) (33)
N-(4-methylphenyl)-ketenimine	[5110-45-2]	yellow crystals	mp 82–83 (34)
triphenylketeneimine	[14181-84-1]	yellow crystals	mp 55–56 (35)
etheneimine	[17619-22-6]	not isolated	
(2-methyl-prop-1-enyliden)aniline	[14016-34-3]	liquid	bp 99–100 (1.6 kPa) (36)

^a At 101.3 kPa unless otherwise stated; to convert kPa to mm Hg, multiply by 7.5.

Simple olefins do not usually add well to ketenes except to ketoketenes and halogenated ketenes. Mild Lewis acids as well as bases often increase the rate of the cycloaddition. The cycloaddition of ketenes to acetylenes yields cyclobutenones. The cycloaddition of ketenes to aldehydes and ketones yields oxetanones. The reaction can also be base-catalyzed if the reactant contains electron-poor carbonyl bonds. Optically active bases lead to chiral lactones (41–43). The dimerization of the ketene itself is the main competing reaction. This process precludes the parent compound ketene from many [2 + 2] cycloadditions. Intramolecular cycloaddition reactions of ketenes are known and have been reviewed (7).

Dimerization. A special case of the [2 + 2] cycloadditions is the dimerization of ketenes. Of the six possible isomeric structures, only the 1,3-cyclobutanediones and the 2-oxetanones (β -lactones) are usually formed. Ketene itself gives predominantly (80–90 %) the lactone dimer, 4-methylene-2-oxetanone (3), called diketene [674-82-8]; approximately 5 % is converted to the symmetrical dimer, 1,3-cyclobutanedione [15506-53-3] (4) which undergoes enol-acetylation to so-called triketene [38425-52-4] (5) (44).

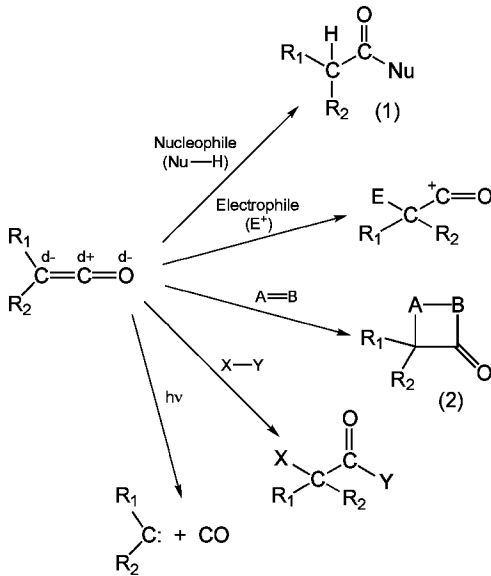
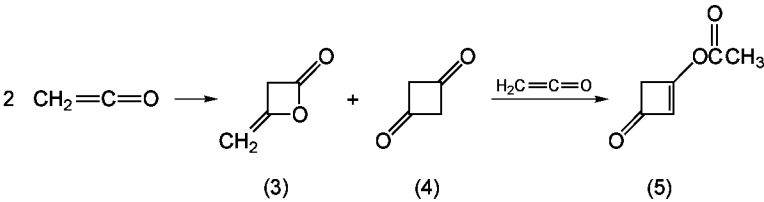
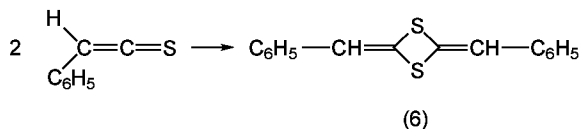


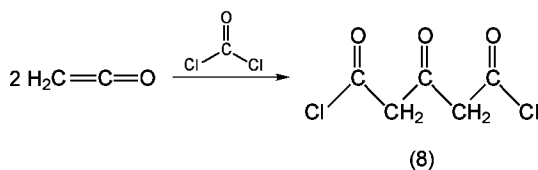
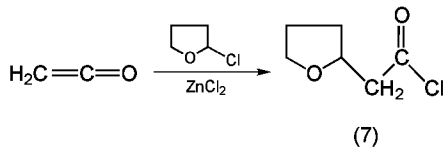
Fig. 1. Reactions of ketenes.

Aldoketenes also form predominantly the lactone dimers, although the ratio of isomers can be influenced by base catalysis. Ketoketenes dimerize symmetrically, and at a slower rate, to 1,3-cyclobutanediones, unless acidic or basic catalysts are present.

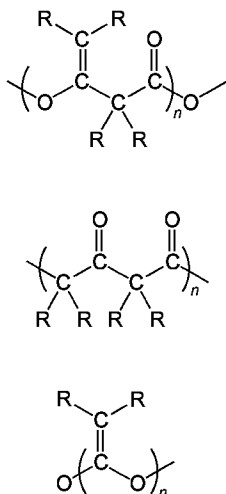
Sterically hindered or very electrophilic substituted ketenes, such as diphenylketene, di-*tert*-butylketene [19824-34-1], and bis(trifluoromethyl)ketene, are quite stable as monomers. Ketenimines tend to polymerize. The dimerization of thioketenes results in 1,3-dithiacyclobutanones (6) (45), a type of dimer not observed with ketenes.



Ketene Insertions. Ketenes insert into strongly polarized or polarizable single bonds, such as reactive carbon–halogen bonds, giving acid halides (7) and into active acid halides giving halides of β -ketoacids (8) (46). Phosgene [77-44-5] (47) and thiophosgene [463-71-8] (48) also react with ketenes.



Other Reactions. The photolysis of ketenes results in carbenes. The polymerization of ketenes has been reviewed (49). It can lead to polyesters and polyketones (50). The polymerization of higher ketenes results in polyacetals depending on catalysts and conditions. Catalysts such as sodium alkoxides (polyesters), aluminum tribromide (polyketones), and tertiary amines (polyacetals) are used. Polymers from $R_2C=C=O$ may be represented as follows.



Ketenes can react in several ways with organometallic compounds and complexes. They can add as ligands to coordinated metals forming stable ketene, ketenyl, and ketenylidene complexes. Ketenes can be inserted into metal-hydride, metal-alkyl, metal-OR, and metal-NR₂ bonds, react with metal-oxide complexes, and with coordinated ligands. This chemistry has been reviewed (9,51).

Manufacture. Ketenes can be considered the internal anhydrides of the corresponding carboxylic acids, and as such can be made by removing a molecule of water from these acids, either directly or indirectly. Numerous methods to convert a carboxylic acid or derivative to the corresponding ketene have been described (1–4).

Commercially and industrially most important, ketene itself, $\text{H}_2\text{C}=\text{C}=\text{O}$, is produced by pyrolysis of acetic acid [64-19-7]. In this process, high quality acetic acid is evaporated and the vapor passed continuously through a radiant coil under reduced pressure at 740–760°C.

The materials of construction of the radiant coil are highly heat-resistant steel alloys, such as Sicromal containing 25° ° Cr, 20° ° Ni, and 2° ° Si. Triethyl phosphate [78-40-0] catalyst is injected into the acetic acid vapor. Ammonia [7664-41-7] is added to the gas mixture leaving the furnace to neutralize the catalyst and thus prevent ketene and water from recombining. The crude ketene obtained from this process contains water, acetic acid, acetic anhydride, and 7 vol ° ° other gases (mainly carbon monoxide [630-08-0], carbon dioxide [124-38-9], ethylene [74-85-1], and methane [74-82-8]). The gas mixture is chilled to less than 100°C to remove water, unconverted acetic acid, and the acetic anhydride formed as a liquid phase (52,53).

The production of ketene by this method has no significant environmental impact. The off-gases from the ketene furnace are either circulated to the furnace and burned to save energy or led to a flare system. The reaction can also be carried out at 350–550°C in the presence of alkali-exchanged zeolite catalysts (54). Small quantities of ketene are prepared by pyrolysis of acetone [67-64-1] at 500–700°C in a commercially available ketene lamp (55,56).

For the two most important industrial uses, the gaseous ketene is immediately treated with acetic acid to form acetic anhydride or dimerized to diketene.

The manufacture of the highly pure ketene required for katenization and acetylation reactions is based on the pyrolysis of diketene, a method which has been employed in industrial manufacture. Conversion of diketene to monomeric ketene is accomplished on an industrial scale by passing diketene vapor through a tube heated to 350–600°C. Thus, a convenient and technically feasible process for producing ketene uncontaminated by methane, other hydrocarbons, and carbon oxides, is available. Based on the feasibility of this process, diketene can be considered a more stable form of the unstable ketene.

Acetone and acetic anhydride can be used as alternative raw materials to acetic acid in the industrial production of ketene. However, the use of

acetone as a raw material is only interesting when the price is low. It is believed that acetone is still used in the former Soviet Union for the production of ketene.

Other methods for the preparation of aldo- and ketoketenes (57) are the pyrolysis or photolysis of ketones, pyrolysis of carboxylic acids and anhydrides, (especially malonic acid mixed anhydrides) (58), thermolysis or photolysis of α -diazoketones via a carbene intermediate and dehalogenation of α -halogen acid chlorides with zinc [7440-66-6] (59). The method of dehydrohalogenation of carboxylic acid chlorides with a base, usually a trialkylamine (60) in producing ketenes, is used for the only other ketenes which are produced industrially, the long-chain (C10–C20) monoalkylketenes (aldoketenes). The starting material in these syntheses are fatty acid chlorides, in particular, stearoyl chloride [112-76-5]. Their lactone dimers have been used for paper sizing for many years. The intermediate, reactive ketenes dimerize to the lactone dimers in the presence of acidic catalysts (61,62).

Thioketenes can be prepared in several ways, from carboxylic acid chlorides by thionation with phosphorus pentasulfide [1314-80-3], P_2S_5 , from ketene dithioacetals by β -elimination, from 1,2,3-thiadiazoles with flash pyrolysis, and from alkynyl sulfides (thioacetylenes). The dimerization of thioketenes to 2,4-bis(alkylidene)-1,3-dithietane compounds occurs quickly. They can be cleaved back pyrolytically (63). For a review see Reference 18.

Ketenimines are usually prepared from carboxylic acid derivatives such as amides and imino chlorides via elimination and from nitriles via alkylation with alkyl halides under strong basic conditions (21,64).

Shipping and Storage. Most ketenes are extremely reactive and unstable so they cannot be stored or transported. Some have been isolated only in solution, or have not been isolated at all, but are used *in situ*. Ketene itself is stable for some hours at -80°C , but dimerizes within minutes at 0°C . It cannot even be transferred through a pipe for any significant distance, even within the same plant, and has to be used *in situ* immediately. All reactions with ketene on an industrial scale have been performed either directly in the ketene manufacturing plant or by transporting diketene and cracking it back to ketene immediately next to the reaction vessel (65). Information on emergency procedures in some safety data sheets (66) give the misleading impression that ketene is stored in cylinders, which is simply not feasible. Some sterically hindered ketoketenes are reasonably stable at room temperature.

Economic Aspects. Due to the physical nature and instability of these materials, all ketene production is used captively and production figures are not readily available. The economic aspects of the products made from ketenes will be addressed later.

Analytical and Test Methods. For a review of detection, determination, and identification of ketenes see Reference 67. Typical properties are the strong ir absorption bands at 2151 cm^{-1} (C—O) and at 1120 cm^{-1} as well as a very low field signal of the sp -hybridized carbon at approximately 194 to 206 ppm and a very high field signal of the sp^2 hybridized carbon at approximately 2.5 to 27 ppm in ^{13}C -nmr spectroscopy.

Health and Safety Factors. Ketene itself is a highly poisonous gas, strongly irritating to the eye, the respiratory tract, and the skin (66,68). Different, sometimes conflicting values for its toxicity are found in the literature (69,70), mainly due to the difficulty in maintaining accurately and measuring low levels of the unstable ketene over hours. Its toxicity is estimated to be of the same order of magnitude as that of phosgene (71), and like the latter it can cause latent damage of the respiratory tract which may become acute only several hours after exposure (pulmonary edema). Repeated or high exposure may cause permanent lung damage.

The LC_{50} (lowest possible lethal concentration) has been reported to be 23 ppm for a 30 min exposure time (mouse), 53 ppm for an exposure time of 100 min (rat, rabbit, and guinea pig), and 200 ppm for an exposure time of 10 min (monkey). No toxic effects were reported upon exposure to 1 ppm for 7 h d over 55 days. The oral LD_{50} (rat) of ketene is 1300 mg/kg, the low level of toxicity probably being due to the almost immediate formation of acetic acid and other acetates in the digestive tract.

The OSHA PEL and NIOSH REL (recommended exposure limit) 10 h-TWA exposure limit for ketene is 0.5 ppm (0.9 mg m^{-3}) (72). This is also the MAK (maximum allowable concentration) value in Germany (73) and Switzerland (74). The NIOSH maximal allowable short-time exposure concentration (TLV-STEL) is 1.5 ppm in the United States, and in Germany 1 ppm for no longer than 5 min and no more than 8 times a day. According to Reference 66, exposure to 25 ppm is immediately dangerous to life and health. Inhalation of even small quantities of ketene leave a characteristic, long lasting unpleasant taste; smokers seem to be particularly sensitive to this. No carcinogenic effects of ketene have been reported. Ketene is listed in the EPA TSCA chemical inventory (1990) and in the EPA TSCA Test Submission (TSCATS) Data Base and the 1992 NHOS Hazard Code 41840.

Practically nothing is known about the toxicity of higher ketenes, thioketenes, and ketenimines, but it is prudent to consider them at least as toxic and hazardous as ketene itself. Ketenes and related compounds are highly reactive with a wide variety of substances. They can polymerize violently or even explosively especially in the presence of bases or strong acids. Ketoketenes are somewhat more stable than aldoketenes but they are prone to oxidation by air, forming highly explosive peroxides (75), and should be handled under nitrogen or argon. Ketene itself is stable in dry air and oxygen [7782-44-7]. Like most volatile organic compounds, most ketenes are flammable and form explosive mixtures with air. Contrary to the statement in Reference 66 under "Emergency Information," ketene itself is a flammable gas which forms explosive mixtures with air in the range 5.5–18 vol %. The liquid has a flash point of -107°C . In case of fire, water or extinguishing powder should be used (76).

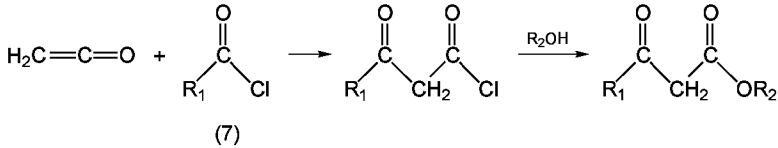
Uses. The lowest member of this class, ketene itself, is a powerful acetylating agent, reacting with compounds containing a labile hydrogen atom to give acetyl derivatives. This reaction is used only when the standard acetylation methods with acetic anhydride or acetyl chloride [75-36-5] do not work well. Most of the ketene produced worldwide is used in the production of acetic anhydride. Acetic anhydride is prepared from the reaction of ketene and acetic acid.

This process is one of the three commercially practiced processes for the production of acetic anhydride. The other two are the oxidation of acetaldehyde [75-07-0] and the carbonylation of methyl acetate [79-20-9] in the presence of a rhodium catalyst (coal gasification technology, Halcon process) (77). The latter process was put into operation by Tennessee Eastman in 1983. In the United States the total acetic anhydride production has been reported to be in the order of 1000 metric tons.

The second most important use of ketene is in the production of diketene [674-82-8] (3) by controlled dimerization. Diketene has wide utility in the manufacture of pharmaceutical and agricultural chemicals, dyes, pigments and fine chemicals.

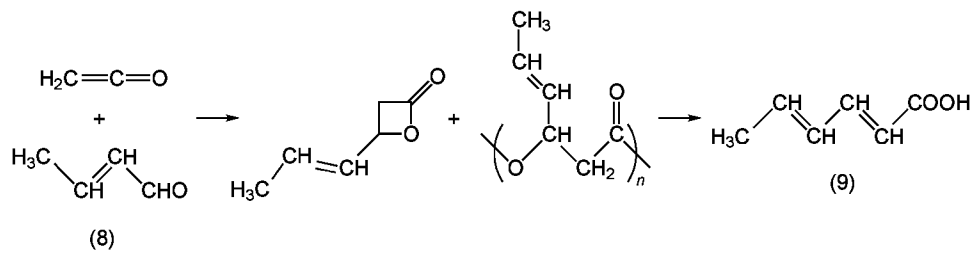
Chlorine adds to ketene to form chloroacetyl chloride [79-04-9] (78). Chloroacetyl chloride (CAC) is used in large volume in the manufacture of the pre-emergence herbicides alachlor [15972-60-8] and butachlor [23184-66-9]. It is estimated that the CAC requirement for this application was in excess of 45,000 metric tons in 1992. Significant volumes of CAC are also used in pharmaceutical manufacture, such as anesthetics of the lidocaine type, and in the production of the tear gas chloroacetophenone [532-27-4]. Other commercial methods for the manufacture of CAC have been described (79).

In the presence of strong acid, such as boron trifluoride [7637-07-2], appropriately substituted acyl chlorides (7, $R_1 = \text{CF}_3, \text{CCl}_3$) add to ketene to form the corresponding acetoacetyl chlorides, which can further react with alcohols to form the corresponding acetoacetates.

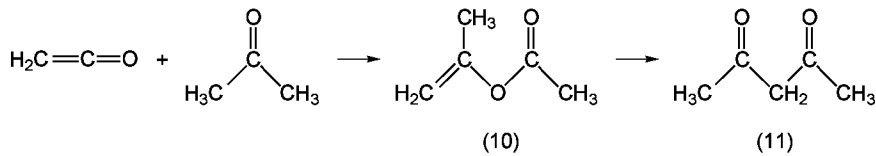


Of industrial significance are ethyl 4,4,4-trifluoroacetoacetate [372-31-6], methyl 4,4,4-trifluoroacetoacetate, and isopropyl 4,4,4-trifluoroacetoacetate for the production of herbicides (eg, Monsanto's Dimension) and antimalarial agents such as Roche's Mefloquin [51773-92-3], as well as ethyl 4,4,4-trichloroacetoacetate [3702-98-5] for the production of pharmaceuticals.

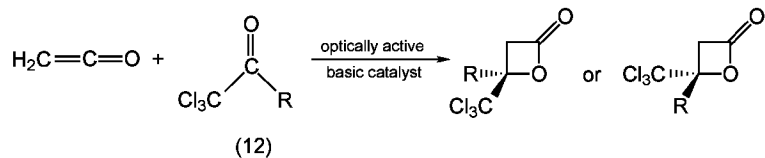
Another principal use of ketene is in the production of sorbic acid [110-44-1] (80,81). In this process, which requires an acidic or manganese(II) catalyst, ketene adds to crotonaldehyde [123-73-9] (8) with subsequent conversion of the β -lactone and the polyester to sorbic acid (qv) (9).



Ketones with labile hydrogen atoms undergo enol acetylation on reaction with ketene. Strong acid catalysis is required. If acetone is used, isopropenyl acetate [108-22-5] (10) is formed (82–85). Isopropenyl acetate is the starting material for the production of 2,4-pentanedione (acetylacetone) [123-54-6] (11).



Ketene can also be added to trihalosubstituted aldehydes or ketones (12) to form 4-trihalomethyloxetanones. If this addition is performed in the presence of optically active bases such as quinine [130-95-0] chiral lactones are obtained (41,42).



Ketene has also been used on a large scale for C-acetylation in the synthesis of the carbapenem antibiotic thienamycin [59995-64-1] (86,87).

Dimeric Ketenes

Physical Properties. Dimeric ketenes are colorless to dark brown liquids or crystalline solids with a broad range of melting and boiling points. Table 2 lists examples of dimeric ketenes and thioketenes.

Table 2. Dimeric Ketenes

Name	CAS Registry Number	Physical properties	Mp or bp, ^a °C
diketene	[674-82-8]	colorless liquid	mp 7.5, bp 127 (101.3 kPa), bp 69–71 (13.3 kPa), bp 38.5 (3.1 kPa)
hexadecylketene dimer	[10126-68-8]	colorless crystals	mp 64 (88),also 81 (89)
dimethylketene dimer	[3173-79-3]	lachrimatory liquid ^b	bp 170 (97.6 kPa) (90), bp 83–85 (5.3 kPa)
octadecylketene dimer	[24430-01-1]	crystals	mp 80 (91)
tetradecylketene dimer	[42272-70-8]	crystals	mp 57–58 (89)
dimethylketene dimer	[933-52-8]	white crystals, sublimates at 95°C	mp 108–111 (92)
dimethylthioketene dimer	[10181-56-3]	crystals	mp 123.5–125 (93)
cyclobutane-1,3-dione	[15506-53-3]	white crystals	mp 119–120 dec (94)
dispiro(5.1.5.1)tetra-decane-7,14-dione	[950-21-0]	crystals	mp 164–165 (95)

^a At 101.3 kPa unless otherwise stated; to convert kPa to torr, multiply by 7.5.

^b Density 1.88 g mL; *n*_D 1.4381.

Chemical Properties. Diketene is a reactive and versatile compound which can undergo reaction with a large variety of compounds. These reactions have been reviewed comprehensively (1,5,6,96).

In most reactions diketene appears to react as acetylketene or one of its tautomeric forms. This is one of the reasons for the correct structure of diketene being firmly established only in 1952, 45 years after its discovery (97,98).

Diketene usually reacts either at the carbonyl group (nucleophilic attack), or at the olefinic bond (electrophilic attack), either process resulting almost always in an exothermic reaction and in the opening of the strained diketene ring (Fig. 2). The strain energy is 94.2 kJ mol (22.5 kcal mol). The so-formed 1,3-dicarbonyl compounds can react further if other functional groups are present often forming heterocyclic compounds. Acetoacetylations and the formation of five- and six-membered heterocyclic rings are the very heart of diketene chemistry.

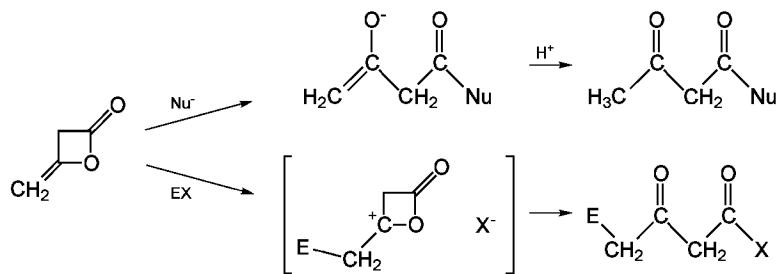


Fig. 2. The reactions of diketene with nucleophiles (Nu[−]) and electrophiles (EX).

It is, however, possible to perform reactions such as hydrogenation, halogenation, polymerization, and [2 + 2] cycloadditions with the exocyclic

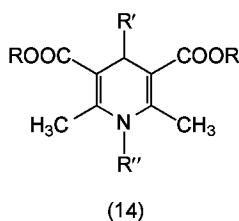
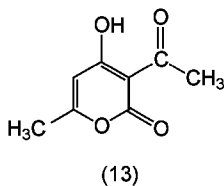
double bond of diketene without opening the β -lactone ring.

Acetoacetylation Reactions. The best known and commercially most important reaction of diketene is the acetoacetylation of nucleophiles to give derivatives of acetoacetic acid (Fig. 2) (1,5,6). A wide variety of substances with acidic hydrogens can be acetoacetylated. This includes alcohols, amines, phenols, thiols, carboxylic acids, amides, ureas, thioureas, urethanes, and sulfonamides. Where more than one functional group is present, ring closure often follows acetoacetylation, giving access to a variety of heterocyclic compounds. These reactions often require catalysts in the form of tertiary amines, acids, and mercury salts. Acetoacetate esters and acetoacetamides are the most important industrial intermediates prepared from diketene.

Diketene is used to C-acetoacetylate aromatic compounds in the presence of aluminum trichloride [7446-70-0]. Benzene [71-43-2] and diketene react to produce acetoacetylbenzene [93-91-4]. Pyrrole [109-97-7] and diketene react to produce 2-acetoacetylpyrrole [22441-25-4]. The C-acetoacetyl derivatives of active methylene compounds such as cyanoacetates, malonodinitrile [109-77-3], and Meldrum's acid [2033-24-1], and olefins can be prepared using diketene.

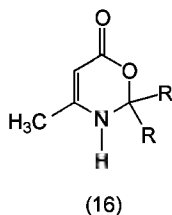
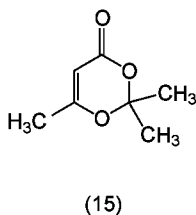
Water hydrolyzes pure diketene only slowly to give acetoacetic acid [541-50-4] which quickly decomposes to acetone and carbon dioxide, but increasing the pH or adding catalysts (amines, palladium compounds) increases the rate of hydrolysis. The solvolysis of diketene in ammonia results in acetoacetamide [5977-14-0] if used in stoichiometric amounts (99), and β -aminocrotonamide [15846-25-0] if used in excess (100).

Six-Membered Heterocycle Ring Formation. Heterocycle formation involving diketene usually involves acetoacetylation of a substrate, followed by intramolecular condensation. Diketene itself readily dimerizes through self-condensation forming mainly dehydroacetic acid [771-03-9] (DHA) (13). Dehydroacetic acid and sodium dehydroacetate [4418-26-2] are used as preservatives for foods and cosmetics. DHA is found as an unwanted by-product in many diketene reactions, but can be obtained intentionally by dimerizing diketene in the presence of pyridine [110-86-1] in benzene, diazabicyclo[2.2.2]octane [280-57-9] (DABCO), and other basic catalysts.



Another important reaction of diketene derivatives is the Hantzsch pyridine synthesis (101). This synthesis is the preparation of 1,4-dihydropyridines (14) starting either from two acetoacetic esters, which react with an aldehyde and ammonia or a primary amine or from 3-aminocrotonates and 2-alkylidene acetoacetic esters, both diketene derivatives. Several such dihydropyridines such as nifedipine [21829-25-4] (102), nimodipine [66085-59-4], and nicardipine [55985-32-5] exhibit interesting pharmaceutical activity as vasodilators (blood vessel dilation) and antihypertensives (see CARDIOVASCULAR AGENTS).

Six-membered heterocycles with two heteroatoms are prepared by reaction of diketene with a substrate containing a C=O or C=N multiple bond. With carbonyl compounds diketene reacts in the presence of acids to give 1,3-dioxin-4-ones. The best known is 2,2,6-trimethyl-4*H*-1,3-dioxin-4-one [5394-63-8] (15), the so-called diketene-acetone adduct, often used as a diketene replacement that is safer to handle and to transport, albeit somewhat less reactive than diketene itself (103,104), forming acetylketene upon heating.



Diketene reacts with imines to give 1,3-oxazinones (16) (105). This reaction has been used in the synthesis of the tranquilizer Ketazolam [27223-35-4] from diazepam [439-14-5] (106).

Other six-membered rings with two heteroatoms are also obtained from reaction of diketene with imidates, cyanamides, carbodiimides, isocyanates, oxazolines, or other multifunctional compounds.

Five-Membered Heterocycle Ring Formation. Diketene reacts with a variety of bifunctional compounds to give five-membered heterocycles, usually through acetoacetylation followed by intramolecular condensation. Thus hydrazines give pyrazolones (107), important dyestuff intermediates, hydrazones give pyrazolin-3-ones, hydroxylamines afford 3-methylisoxazol-5-ones or isomers depending on conditions (108), hydroxamic acids give oxazoles, α -hydroxyketones can give butenolides (109), and α -hydroxy acids give furanones (110). Several of these five-membered heterocycles are intermediates for pharmaceuticals and agrochemicals. A variety of five- and six-membered heterocycles are also obtained indirectly from diketene via 2- and 4-chloroacetoacetates.

Other Reactions of Diketene. Diketene reacts with elemental chlorine to give 4-chloroacetoacetyl chloride [41295-64-1], which can further react to the corresponding esters (111).

4-Chloroacetoacetic esters are important industrial intermediates used especially for the synthesis of the aminothiazolylacetic acid side chain of modern cephalosporins (see ANTIBIOTICS, β -LACTAMS, CEPHALOSPORINS). For a review of the chemistry of 4-chloroacetoacetates see Reference 112.

Diketene can be hydrogenated to β -butyrolactone [3068-88-0]. In the presence of 2,2'-bis(diphenylphosphino)-1,1'-binaphthylruthenium(II) complexes the reaction is stereoselective (113).

Ozonolysis of diketene affords formaldehyde [50-00-0] and the very un-stable malonic anhydride [15159-48-5] (114), which readily decomposes to CO₂ and ketene, but can be converted at low temperature to malonic acid monoesters or Meldrum's acid (115) (see MALONIC ACID AND DERIVATIVES).

Pyrolysis of diketene at temperatures greater than 400°C gives two molecules of ketene. This method has been used industrially. At present there is no method to convert diketene efficiently into allene [463-49-0] and CO₂, the thermodynamic products.

Grignard reagents add to diketene in the presence of cobalt iodide [15238-00-3], CoI₂, or palladium [440-05-3] to give 3-methylenecarboxylic acids, used in terpenoid and hormone syntheses, as well as monomers for radical copolymers (116,117) (see HORMONES; TERPENOIDS).

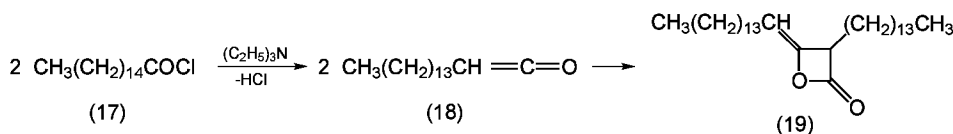
Diketene also reacts with organometallic compounds including organoboron, organosilicon, organoaluminum, and organotin compounds to afford acetoacetate complexes (118). Diketene can be polymerized with various catalysts such as boron trifluoride, mercury(II) chloride [7847-94-7], HgCl₂, and ion-exchange resins to low molecular weight polymers containing unconjugated methylene groups (119,120). Diketene can also be copolymerized under radical conditions with monomers such as vinyl chloride [75-01-4], vinyl acetate [108-05-4], and acrylonitrile [107-13-1], giving polymers containing the β-lactone ring in the monomer units (121). The lactone ring can further react with nucleophiles to produce elaborately functionalized polymers.

Dimeric aldoketenes and ketoketenes of β-lactone structure show a chemical behavior which is not much different to that of diketene. Thus nucleophiles add in similar fashion to give derivatives of 3-ketoacids which are mono- or dialkylated at C-2 (aldo- and ketoketene dimers, respectively), but the reaction can often be slower than with the parent compound and, in case of long-chain or bulky substituents, may not proceed at all. Other reactions can proceed differently than those with diketene. For an overview of important reactions of aldoketene and ketoketene dimers see Reference 122.

Manufacture. Of the industrially used ketene dimers, the first member of the group, diketene itself, is by far the most significant. The only other ketene dimers used on an industrial scale are the long-chain alkylketene dimers (AKDs) (usually C-12 to C-20) which are made by Hercules for paper sizing applications. Diketene is made industrially by controlled dimerization of ketene. The gaseous ketene from the ketene furnace, dried and cooled and at a pressure of 20–30 kPa before entering the dimerization unit, is dimerized in a liquid ring pump. The dimerization is preferably carried out in diketene itself (52,123), and gives yields up to 85% based on acetic acid.

Crude diketene obtained from the dimerization of ketene is dark brown and contains up to 10% higher ketene oligomers but can be used without further purification. In the crude form, however, diketene has only limited stability. Therefore, especially if it has to be stored for some time, the crude diketene is distilled to >99.5% purity (124). The tarry distillation residue, containing triketene (5) and other oligomers, tends to undergo violent spontaneous decomposition and is neutralized immediately with water or a low alcohol. Ultrapure diketene (99.99%) can be obtained by crystallization (125,126). Diketene can be stabilized to some extent with agents such as alcohols and even small quantities of water [7732-18-5] (127), phenols, boron oxides, sulfur [7704-34-9] (128) and sulfate salts, eg, anhydrous copper sulfate [7758-98-7].

Long-chain alkylketene dimers are also produced from their corresponding monomers by controlled dimerization (58,59). The starting materials are mixed fatty acids (C₁₂–C₂₂, mostly C₁–C₂₀), which are converted to the acid chlorides (17) using phosphorus trichloride [7719-12-2], and then to the ketenes (18) through the removal of hydrogen chloride [7647-01-0] with triethylamine [121-44-8]. These intermediary long-chain aldoketenes combine to the unsymmetrical dimers (19) (β-lactones), with a dione dimer content of <1%. The raw product of AKDs is 80–85% pure as β-lactone. The rest is free acids, their anhydrides, and some pyrones.



This purity is adequate for conversion into aqueous emulsions which are supplied to the paper industry.

Shipping and Storage. Because of its extreme reactivity and hazardous properties, diketene is now generally consumed at the site of production. Regulations still allow the transport of diketene by road and ship (transport by plane has recently been forbidden) but the two largest producers in the United States, Lonza Inc. and Tennessee Eastman, several years ago voluntarily agreed to no longer ship diketene. Diketene, however, was shipped in refrigerated tank cars for many years without incident when it was still sold on the merchant market (129).

Pure diketene is stable for several weeks if stored at or below 0°C in an aluminum or stainless steel container. Glass should be avoided because of its inherent basicity which favors slow polymerization. Above 15°C slow decomposition occurs and the color becomes progressively darker. Pressure build-up upon prolonged exposure to heat is possible. Heating and contamination of the container, especially by acids, bases, and water, should be avoided. Residual vapors in empty containers are hazardous and may explode on ignition.

A shippable but somewhat less reactive form of diketene is its acetone adduct, 2,2,6-trimethyl-4*H*-1,3-dioxin-4-one (15) (103,104). Thermolysis of this safer to handle compound provides acetylketene, a reactive intermediate that can be used for acetoacetylation and cycloaddition reactions. The diketene–acetone adduct as well as *tert*-butylacetoacetate [1694-31-1] (also used for acetoacetylations by the transacetoacetylation reaction) (130), are offered commercially.

The higher, long-chain dimers as well as the tetramer dehydroacetic acid are far more stable and can be handled safely. The alkylketene dimers (AKDs) are shipped to the paper industry in tank trucks in the form of ready-to-use aqueous emulsions with a total solids content of 12–25% and a guaranteed shelf life of 30 days, as they have good hydrolytic stability. In this form they are not combustible liquids, and are listed in the Canadian Domestic Substances List.

Economic Aspects. All diketene production is used captively and, therefore, production figures can only be estimated by the volume of derivatives output on the merchant market. World production of diketene is probably close to 100,000 metric tons, approximately 20% of that production in the United States. Before the discontinuation of all shipment in the United States, diketene was sold for less than \$2.20/kg.

The world production of the alkylketene dimers is believed to be around 15,000 metric tons, equally split between the United States and the rest of the world. The value of the AKDs (80–85% assay) is \$4.00–4.40/kg.

Analytical and Test Methods. Colorimetric qualitative tests for diketene are known but seldom used (131). Identification is by spectrometric methods. Diketene has typical ir absorption bands at ~1880, 1855, and 1685 cm⁻¹, and signals at 3.92 (t), 4.51 (m), and 4.87 (m) ppm in the ¹H-nmr spectrum (CDCl₃). Purity is routinely monitored by gc. Alternatively, diketene is quantitatively converted to acetoacetic derivatives which are assayed by standard methods.

Health and Safety Factors

Good ventilation, eye and skin protection, and an approved organic vapor respirator should be used when handling diketene.

Diketene is a strongly irritating, powerfully lachrimatory, poisonous liquid, but is considerably less toxic than ketene (68). The eye and respiratory tract are especially endangered, as diketene quickly damages the tissue of the cornea and lung. Lung edema may occur even up to two days after inhalation. Exposure causes a burning sensation in eyes, nose, and throat, as well as respiration difficulties and coughing. At higher levels, loss of consciousness and death can occur. Absorption of liquid diketene by the skin is possible, with local itching and severe burning of the skin (131). Ingestion causes irritation of the gastrointestinal tract.

Chronic effects are not known and diketene was not found to be carcinogenic on skin application, injection beneath the skin, or gastric feeding in mice and rats (132–134). It is not listed as a carcinogen by International Agency for Research on Cancer (IARC), the National Toxicology Program (U.S.), OSHA, and ACGIH.

Since diketene is a strong eye irritant even at low levels, it has a strong warning effect. Diketene becomes unbearable before acute toxic levels are reached. Due to the risk of delayed lung edema, a physician should be consulted and the patient monitored carefully after exposure.

Rat LD₅₀ is 560 mg kg (rat) and the lowest lethal dose, LD₁₀ is 800 mg kg (mouse). By inhalation, the LC₅₀ is 3 g m⁻³ for a 2 h period which is approximately 860 ppm (guinea pig) and the LC₁₀ is 20,000 ppm 1 h (rat). By skin application, the LD₅₀ is 2830 mg kg (rabbit). The DOT skin irritation index is 5.0 (strongly irritating). No ACGIH threshold limit values or OSHA permissible exposure limits have been established. No odor threshold is available.

Very little is known about the toxicology of other dimeric ketenes. For the dimeric dimethylketene there is equivocal evidence of tumors resulting from massive exposure in rats reported for the β-lactone form (3,3-dimethyl-4-isopropylidene-2-oxetanone), whereas the symmetric form (2,2,4,4-tetramethylcyclobutane-1,3-dione) induces tumors in mice after lengthy skin applications.

Diketene is a flammable liquid with a flash point of 33°C and an autoignition temperature of 275°C. It decomposes rapidly above 98°C with slow decomposition occurring even at RT. The vapors are denser than air (relative density 2.9, air air = 1). The explosive limits in air are 2–11.7 vol % (135). In case of fire, water mist, light and stabilized foam, as well as powder of the potassium or ammonium sulfate-type should be used. Do not use basic extinguisher powders and do not add water to a closed container.

The greatest hazard is violent exothermic polymerization with quick pressure build-up and rupture of the vessel. Pressure increases of 12.1 MPa min have been measured. Polymerization can take place on heating or through contact with even catalytic amounts of bases, mineral acids, strong oxidants, Friedels-Crafts catalysts, and other substances. On addition of 1 drop of an amine to 1 g of diketene, the hot reaction mixture is violently projected out of the test tube within a few seconds. Water decomposes diketene slowly. Acids react slower than bases; storage vessels should therefore be thoroughly clean and free of contaminants. In chemical reactions diketene should be added slowly to the other reagents, not vice-versa, making sure that there is no build-up of unreacted diketene.

Higher dimeric ketenes are flammable but have higher flash points and are less reactive than diketene. Almost no data are available. Diketene can be disposed of by incineration, preferably after dilution with an inert solvent such as toluene. Higher ketene dimers can also be incinerated.

Uses. Diketene is described as having bactericidal activity (136,137). It also raises the octane rating of gasoline (138) but is not used for these purposes in the Western world.

As the most reactive and economical source of the acetoacetyl moiety, diketene is used as a valuable synthetic intermediate in the manufacture of acetoacetic acid derivatives and heterocyclic compounds which are used as intermediates in the manufacture of dyestuffs, agrochemicals, pharmaceuticals, and polymers.

The AKDs are used in paper sizing applications. Paper sizing chemicals provide paper and paperboard with resistance to wetting by liquids, especially water repellency in paper cups, milk cartons, photographic paper, coatings, and packaging paper (139) (see PAPER; PAPERMAKING ADDITIVES).

Ketene trimer can be recovered from the tarry residue of diketene distillation and converted into valuable building blocks like 1,3-cyclobutanedione and squaric acid [2892-51-5] (140,141), an important intermediate in the synthesis of pharmaceuticals and squarylium dyes used in photostatic reproduction (142,143).

Acetoacetic Acid Derivatives

The most important use of diketene is for the preparation of derivatives of acetoacetic acid, such as acetoacetate esters, acetoacetamides, and chloroacetoacetates, which have found many uses in life sciences, dyestuffs, adhesives, and coatings.

Physical Properties. Acetoacetic esters are high boiling liquids with pleasant odors. Lower N-alkylamides are water-soluble liquids; acetoacetamide and acetoacetarylides are solids. 4-Chloroacetoacetates are high boiling lachrymatory liquids. Some physical properties are listed in Tables 3 and 4.

Table 3. First Generation Diketene Derivatives

Diketene derivative	CAS Registry Number	Mp or bp, ^a °C	Principal uses
methyl acetoacetate	[105-45-3]	bp 169–170, bp 69–71 (1.8 kPa)	herbicides, pesticides, chemical synthesis, protection of NH ₂ groups
ethyl acetoacetate	[141-97-9]	bp 181, bp 74 (1.9 kPa)	antibiotics, protection of NH ₂ groups, chemical synthesis
tert-butyl acetoacetate	[1694-31-1]	bp 76–78 (2.0 kPa)	acetoacetylations
2-acetoacetoxyethyl methacrylate	[21282-97-3]	bp 100 (0.1 kPa)	acetoacetylated polymers, coatings
methyl 4-chloro- acetoacetate	[32807-28-6]	mp 14, bp 50 (0.133 kPa)	L-carnitine amino-thiazole–acetic acid derivatives
ethyl 4-chloro- acetoacetate	[638-07-3]	mp 8.5, bp 95 (1.33 kPa)	coumarins, aminothiazole–acetic acid deriva-tives
acetoacetamide	[5977-14-0]	mp 54	methylpyrazolones
N-methyl- acetoacetamide	[20306-75-6]	bp 96–98 (13 kPa)	insecticides, polyester curing
N,N-dimethyl-acetoacetamide	[2044-64-6]	bp 75–77 (0.2 kPa)	insecticides, polyester curing
N-(β-hydroxy)ethyl- acetoacetamide	[24309-97-5]	mp 35–40	animal feed additive
diethylacetoacetamide	[2235-46-3]	bp 123–124 (1.2 kPa)	insecticides
acetoacetanilide	[101-01-2]	mp 84–85	yellow pigments, fungicides
o-acetoacetotoluidide	[93-68-5]	mp 104	orange pigments
o-acetoacetoaniside	[92-15-9]	mp 87	yellow dyes
N-(2,4-dimethyl-phenyl)-3-oxo-butyr amide	[97-36-9]	mp 89	orange and yellow pigments
N-acetoacet- <i>p</i> -phenetidine ^b	[122-82-7]	mp 103.5	yellow road pigment
dehydroacetic acid	[520-45-6]	mp 109–111	antimicrobials
diketene–acetone adduct	[5394-63-8]	bp 65–67 (15 kPa)	diketene substitute
6-methyluracil	[626-48-2]	mp 270–280 (dec)	dipyridamole

^a At 101.3 kPa, unless otherwise stated; to convert kPa to torr, multiply by 7.5.

^b 4-Ethoxyacetoacteanilide

Table 4. Second Generation Diketene Derivatives

Diketene derivative	CAS Registry Number	Mp or bp, ^a °C	Principal uses
1-phenyl-3-methyl-5-pyrazolone	[89-25-8]	mp 127.5	yellow, orange, and red pigments
1-(<i>p</i> -tolyl)-3-methyl-5-pyrazolone	[86-92-0]	mp 133	orange and red pigments

3-methyl-5-pyrazolone	[108-26-9]	mp 217	yellow and orange dyes
4-hydroxy-6-methyl-pyrone	[675-10-5]	mp 185–187	chemical synthesis
5-aminoorotic acid	[7164-43-4]	mp 250 dec	dipyridamole
ethyl 2-chloroacetoacetate	[609-15-4]	bp 107 (1.87 kPa)	chemical synthesis, imidazoles, vitamin B
methyl 2-chloro-acetoacetate	[4755-81-1]	bp 137	chemical synthesis, imidazoles, vitamin B
methyl 3-aminocrotonate	[14205-39-1]	mp 81–84, bp 112 (55 kPa)	dihydropyridine calcium-channel blockers
methyl 3-oxopentanoate	[30414-53-0]	bp 180.7	ethyl substituted pyrimidines, non-steroidal anti-inflammatories
ethyl 2-(2-aminothiazole-4-yl)-acetate	[53266-94-7]	mp 92–94	side chains of third generation cephalosporins
L-carnitine	[541-15-1]	mp 210–212 dec	dietary supplement
1,4-cyclohexanedione	[637-88-7]	mp 76–79	chemical building block
tetronic acid	[4791-56-6]	mp 141–143 dec	chemical building block, vitamins

^a At 101.3 kPa, unless otherwise stated; to convert kPa to torr, multiply by 7.5.

Chemical Properties. The acetoacetyl moiety is highly functionalized and can undergo many transformations. At the active methylene group, condensation reactions with other carbonyl compounds, halogenation, alkylation, and nitrosation can occur. At the ester or amide group, decarboxylation and trans-esterification can occur. At the ketone carbonyl group, reduction and addition of nucleophiles can occur. Combination of such reactions gives access to a broad spectrum of different types of compounds. For a general overview on the chemistry of acetoacetates see References 96, 144–146; of 4-haloacetoacetates see Reference 112; of acetoacetyrlides and pyrazolones in azo dyes and pigment uses see AZO DYES.

Manufacture and Uses. Acetoacetic esters are generally made from diketene and the corresponding alcohol as a solvent in the presence of a catalyst. In the case of liquid alcohols, manufacturing is carried out by continuous reaction in a tubular reactor with carefully adjusted feeds of diketene, alcohol, and catalyst, or alcohol–catalyst blend followed by continuous purification (Fig. 3). For solid alcohols, an inert solvent is used. Catalysts used include strong acids, tertiary amines, salts such as sodium acetate [127-09-3], organophosphorus compounds, and organometallic compounds (5).

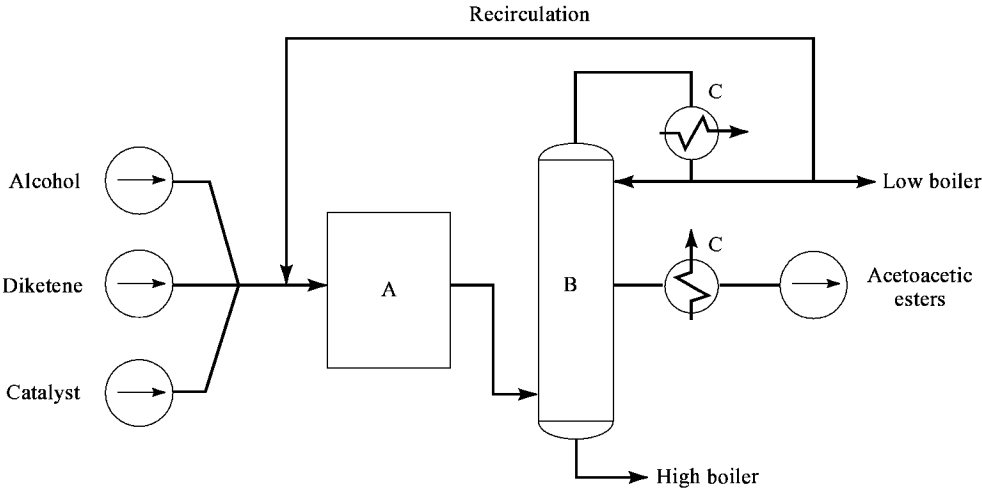


Fig. 3. Flow sheet for the preparation of acetoacetic esters (Lonza). A, reactor; B, rectification column; C, condensor.

Methyl acetoacetate (MAA) and ethyl acetoacetate (EAA) are the most widely used esters; they are found in the pharmaceutical, agricultural, and allied industries. Both esters are used extensively as amine protecting agents in the manufacture of antibiotics and synthetic sweeteners (Dane Salts) (147). Principal outlets for MAA are the manufacture of the organophosphorus insecticide diazinon [333-41-5] (148,149) and the uracil herbicides bromacil [314-40-9] and terbacil [5902-51-2] (150,151) (see INSECT CONTROL TECHNOLOGY; HERBICIDES).

Hydroxyalkyl acrylates and polyols are acetoacetylated with diketene to give comonomers used in adhesives, polymers, and coatings, especially the new low solvent coatings, and for emulsion polymerization. The most widely used compound is 2-acetoacetoxyethyl methacrylate (AAEMA) (152).

Alkylacetoacetamides are often made by reaction of diketene and the corresponding alkylamines in aqueous solutions. They are traded mostly as aqueous solutions, which in the case of the *N*-monoalkyl derivatives have limited stability and require refrigeration for prolonged storage. The acetoacetamides produced from small-chain aliphatic amines are used in the manufacture of systemic insecticides such as monocrotophos (Azodrin) [6923-22-4] (153) and dicrotophos (Bidrin) [141-66-2] (153), phosphamidon (Dimecron) [13171-21-6] (154), and oxamyl (Vydate) [23135-22-0]. With the exception of oxamyl, these insecticides are no longer produced in the United States (see INSECT CONTROL TECHNOLOGY). *N*-Hydroxyethylacetoacetamide [24309-97-5] is used in the manufacture of the animal feed additive olaquindox [23696-28-8] (155) (see FEEDS AND FEED ADDITIVES). Lower aliphatic acetoacetamides are also used as copromoters in the curing of unsaturated polyesters and alkyd coatings (156).

The acetoacetyrlides, are produced in a similar way from diketene and aromatic amines, either in water, butyl acetate [123-86-4] or water–organic solvent mixture. These derivatives are widely used in dyes and pigments. These applications involve the coupling of the acetoacetyrlide with a diazonium salt (Japp-Klingemann reaction) to produce the corresponding diazo compound. CI Pigment Yellow 12 [6358-85-6], one of the largest volume pigments sold in the United States, is produced in this fashion by reaction of acetoacetanilide [102-01-2] (AAA) and the bis(diazo) salt of 3,3'-dichlorobenzidine. Other color pigments are produced from differently substituted arylides and differently substituted coupling components (see AZO DYES). AAA itself is also used in the manufacture of carboxin [5234-68-4], a widely used seed treatment (157).

Shipping and Storage

MAA and EAA are stable liquids, and are shipped in nonreturnable 208-L (55-gal) polyethylene-lined drums. For bulk shipments, insulated stainless steel tank containers and trucks provide secure protection. 2-Acetoacetoxyethyl methacrylate is a liquid stabilized with radical inhibitors such as BHT [128-37-0] and has a shelf life of approximately three months. Shipment is in 60- or 208-L polyethylene-lined drums. Acetoacetyrlides are nicely crystalline, stable solids and are shipped in 208-L drums with polyethylene liners.

Economic Aspects

Total U.S. annual production of MAA and EAA combined is estimated to be 6000–7000 metric tons. The list prices at the end of 1992 for large volumes were \$2.75 kg for MAA and \$3.00 kg for EAA. There are only two U.S. producers of these esters at this time, Tennessee Eastman Co. in Kingsport, Tennessee, and Lonza Inc. in Bayport, Texas.

Total U.S. annual production of all arylides combined is estimated to be 12,000–13,000 metric tons. The largest volume arylide is AAA (acetoacetanilide) for Pigment Yellow 12 as well as for carboxin. The list price of AAA at the end of 1992 was \$3.40 kg.

BIBLIOGRAPHY

"Ketenes" in *ECT* 1st ed., Vol. 8, pp. 109–113, by J. R. Caldwell, Tennessee Eastman Co.; "Ketenes and Related Substances" in *ECT* 2nd ed., Vol. 12, pp. 87–100, by R. H. Hasek, Tennessee Eastman Co.; in *ECT* 3rd ed., Vol. 13, pp. 874–893, by R. H. Hasek, Tennessee Eastman Co.

1. D. Borrmann, in *Houben Weyl Methoden der Organischen Chemie*, Vol. VII 4, 4th ed., Thieme Verlag, Stuttgart, Germany, 1968, pp. 53–447.
2. H. Kropf and E. Schaumann, eds., in *Houben Weyl Methoden der Organischen Chemie*, Vol. E 15, 4th ed., parts 2 and 3, Thieme Verlag, Stuttgart, Germany, 1993, pp. 1598–3146.
3. S. Patai, ed., *The Chemistry of Ketenes, Allenes and Related Compounds*, parts 1 and 2, John Wiley & Sons, Inc., New York, 1980.
4. P. W. Reynolds, in V. H. Agreda and J. R. Zoeller, eds., *Acetic Acid and its Derivatives*, Marcel Dekker, Inc., New York, 1993, p. 161.
5. R. J. Clemens and J. S. Witzeman, in Ref. 4, p. 173.
6. R. J. Clemens, *Chem. Rev.* **86**, 241–318 (1986).
7. B. Snider, *Chem. Rev.* **88**, 793–811 (1988).
8. H. R. Seikaly and T. T. Tidwell, *Tetrahedron* **42**, 2587 (1986).
9. G. L. Geoffroy and S. L. Bassner, in F. G. A. Stone and R. West, eds., *Advances in Organometallic Chemistry*, Vol. 28, Academic Press, Inc., San Diego, Calif., 1988, pp. 1–83.
10. H. W. Moore and O. H. W. Decker, *Chem. Rev.* **86**, 821–830 (1986).
11. R. F. C. Brown and F. W. Eastwood, in Ref. 3, pp. 757–778.
12. C. Wentrup and co-workers, *J. Am. Chem. Soc.* **113**, 3130 (1991).
13. Ref. 2, pp. 3119–3146.
14. T. Kappe and E. Ziegler, *Angew. Chem.* **86**, 529 (1974); *Int. Ed.*, 491 (1974).
15. Z. Cinar and Y. Inel, *Kim Sanay* **31**, 49 (1989).
16. G. Paiaro and L. Pandolfo, *Comments Inorg. Chem.* **12**, 213 (1991).
17. Ref. 2, p. 2530.
18. E. Schaumann, *Tetrahedron* **44**, 1827 (1988).
19. Ref. 2, p. 2531.
20. Ref. 1, pp. 323–339.
21. M. W. Barker and W. E. McHenry in Ref. 3, p. 701.
22. F. O. Rice and co-workers, *J. Amer. Chem. Soc.* **56**, 1760 (1934).
23. D. Darling and R. L. Kidwell, *J. Org. Chem.* **33**, 3974 (1968).
24. L. A. McDougall and J. E. Kilpatrick, *J. Chem. Phys.* **42**, 2311 (1965).
25. M. J. Edwards and J. M. Williams, *J. Chem. Soc.*, 855 (1927).
26. H. Staudinger and E. Ott, *Chem. Ber.* **41**, 2208 (1908).
27. I. L. Knunyants, Y. A. Cherbukov, and M. D. Bargamova, *Izv. Akad. Nauk. SSSR Ser. Khim.*, 1389 (1963).
28. P. G. Blake and K. J. Hole, *J. Phys. Chem.* **70**, 1464 (1966).
29. E. Fuchs and co-workers, *Synthesis*, 1099 (1991).
30. R. A. Ruden, *J. Org. Chem.* **39**, 3607 (1974).
31. H. Pracejus and G. Wallura, *J. Prakt. Chem.* **19**, 33 (1963).
32. H. Bock and co-workers, *J. Am. Chem. Soc.* **99**, 1664 (1977).
33. E. Schaumann and W. Walter, *Chem. Ber.* **107**, 3562 (1974).
34. J. M. Bohen and M. M. Jouillie, *J. Org. Chem.* **38**, 2652 (1973).
35. H. Staudinger and J. Meyer, *Chem. Ber.* **53**, 72 (1920).
36. H. J. Bestmann, J. Lienert, and L. Mott, *Justus Liebigs Ann. Chem.* **718**, 24 (1968).
37. H. Ulrich, *Cycloaddition Reactions of Heterocumulenes*, Academic Press, Inc., New York, 1967, pp. 89–94.
38. D. P. N. Satchell and R. S. Satchell, *Chem. Soc. Rev.* **4**, 231 (1975).
39. H. W. Coower and R. C. Hart, *Chem. Eng. Proc.* **78**, 72 (1982).
40. J. March, *Advanced Organic Chemistry*, 3rd ed., John Wiley & Sons, Inc., New York 1985, p. 762.
41. H. Wynberg and E. G. J. Staring, *J. Am. Chem. Soc.* **104**, 166 (1982), and *J. Org. Chem.* **50**, 1977 (1985).
42. World Pat. Appl. 84 01 577 (Apr. 26, 1984), H. Wynberg and E. G. J. Staring (to the University of Groningen, the Netherlands).
43. *Chiral Building Blocks*, technical brochure, Lonza AG, Basel, Switzerland, 1991.
44. L. Tenud, M. Weilenmann, and E. Dallwigk, *Helv. Chim. Acta* **60**, 975 (1977).
45. G. Purello and P. Fiandaca, *J. Chem. Soc., Perkin Trans. 1*, 692 (1976).
46. Ref. 2, p. 2444.
47. **Swiss Pat. 543,474** (Sept. 15, 1970), F. Broussard (to Lonza AG).
48. J. C. Martin, P. G. Gott, and P. W. Reynolds, *J. Org. Chem.* **46**, 3911 (1981).
49. G. F. Pregaglia and M. Binaghi, in N. M. Bikales, ed. *Encyclopedia of Polymer Science and Technology*, Vol. 8, Wiley-Interscience, New York, 1968, pp. 45–57.
50. K. C. Khemani and F. Wudl, *J. Am. Chem. Soc.* **111**, 9124 (1989).
51. Ref. 2, pp. 2445, 2455, 2527.
52. **Ger. Pat. 687,065** (Dec. 28, 1939), J. Sixt and M. Mugdan (to Consortium für Elektrochem. Ind.).
53. **Ger. Pat. 734,439** (1934), (to Consortium für Elektrochem. Ind.).
54. L. M. Parker, D. M. Bibby, and I. J. Miller, *J. Catal.* **129**, 438 (1991).
55. Technical data, Fischer Labor- und Verfahrenstechnik, Industriepark Kottenforst, D-53340 Meckenheim b. Bonn, Germany.
56. H. Stage, *Chem. Ztg.* **97**, 67 (1973).
57. Ref. 1, pp. 68–108.
58. W. E. Hanford and J. C. Sauer, *Org. Reactions* **3**, 108 (1946).
59. C. W. Smith and D. G. Norton, *Org. Synth.*, Coll Vol. IV, 348 (1963).
60. J. C. Sauer, *Org. Synth.*, Coll Vol. IV, 560 (1963).
61. **Ger. Pat. 885,834** (1937), (to Deutsche Hydrierwerke).
62. *Chem. Eng. News* **33**, 1018 (1955).
63. G. Seybold and C. Heibl, *Chem. Ber.* **110**, 1225 (1977).

64. Ref. 2, p. 2531.
65. Lonza AG, Visp, Switzerland, unpublished results, 1985–1992.
66. *Hazardous Substance Fact Sheet Nr. 1092*, Right-to-Know Project, New Jersey Department of Health, Trenton, N.J., 1987.
67. J. W. Munson, in Ref. 3, pp. 165–188.
68. RTECS, *Registry of Toxic Effects of Chemical Substances*, Vol. 2, NIOSH, Cincinnati, Ohio, 1981–1982, entry OA7700000, p. 595, and RQ8225000, p. 930; also available on CD-ROM through Canadian Centre for Occupational Health and Safety, Hamilton, Ontario, Canada, 1992.
69. K. Verschueren, *Handbook of Environmental Data on Organic Chemicals*, Van Nostrand, New York, 1977, pp. 258, 415.
70. J. J. Gomer, *Zentralbl. Arbeitsmed. Arbeitsschutz* **1**, 37–39 and 58–63 (1951).
71. H. A. Wooster, C. C. Luskbaugh, and C. E. Redemann, *Ind. Hyg. Toxicol.* **29**, 56 (1947).
72. *Documentation of the Threshold Limit Values and Biological Exposure Index*, 5th ed., American Conference of Governmental Industrial Hygienists, Cincinnati, Ohio, 1986.
73. *Maximale Arbeitsplatzkonzentration und biologische Arbeitsstofftoleranzwerte*, VCH Verlagsgesellschaft, Weinheim, Germany, 1987.
74. *Zulässige Werte am Arbeitsplatz*, Schweiz, Unfall Versicherungsanstalt, Luzern, Switzerland, 1987.
75. H. Staudinger, *Die Ketene*, F. Enke Verlag, Stuttgart, Germany, 1912, p. 49.
76. Internal Data Sheet Nr. P0152, Lonza AG, Visp, Switzerland, 1978.
77. **Ger. Pat. 2,441,502** (May 7, 1975), C. Hewlett (to Halcon International).
78. **Ger. Pat. 2,247,764** (Apr. 12, 1973) and 2,247,765 (May 03, 1973), V. Gash, D. Bissing, and E. Donald (to Monsanto).
79. Ref. 4, p. 239.
80. **Ger. Pat. 1,913,097** (Sept. 24, 1970), G. Kuenstle and H. Spes (to Wacker).
81. **U.S. Pat. 3,574,728** (Apr. 13, 1971), I. Takasu, M. Higuchi, and Y. Hijioka (to Daicell).
82. H. Spes, *Chem. Ing. Tech.* **38**, 955 (1966).
83. B. H. Gwynn and E. F. Degering, *J. Am. Chem. Soc.* **64**, 2216 (1942).
84. H. J. Hagemeyer and D. C. Hull, *Ind. Eng. Chem.* **41**, 2920 (1949).
85. R. N. Lacey, *Adv. Org. Chem.* **2**, 213 (1960).
86. D. G. Melillo and co-workers, *J. Org. Chem.* **51**, 1498 (1986).
87. *The Merck Index*, 11th ed., monograph #9240, Merck, Inc., Rahway, N.J., 1989.
88. E. S. Rothman, *J. Org. Chem.* **32**, 1683 (1967).
89. F. Takei, K. Murai, and G. Akazome, *Kogyo Kagaku Zasshi* **60**, 1271 (1957).
90. K. D. Berlin, and R. B. Hanson, *J. Org. Chem.* **32**, 1763 (1967).
91. **Rus. Pat. 59,228,661** (Dec. 22, 1984), (to Kao Corp.).
92. C. C. McCarney, R. S. Ward, and D. W. Roberts, *Tetrahedron* **32**, 1189 (1976).
93. E. U. Elam and H. E. Davis, *J. Org. Chem.* **32**, 1562 (1967).
94. H. H. Wasserman, J. U. Piper, and E. V. Dehmlov, *J. Org. Chem.* **38**, 1451 (1973).
95. K. Kimura and co-workers, *Bull. Chem. Soc. Jpn.* **49**, 741 (1976).
96. A. B. Boese, Jr., *Ind. Chem. Eng.* **32**, 16 (1940).
97. C. D. Hurd and C. A. Blanchard, *J. Am. Chem. Soc.* **72**, 1461 (1950).
98. L. Katz and W. N. Lipscomb, *J. Org. Chem.* **17**, 515 (1952).
99. **Ger. Pat. 3,101,650** (Aug. 26, 1982), G. Kuenstle and H. Jung (to Wacker).
100. **Ger. Pat. 2,037,388** (Feb. 18, 1971), T. Inoi and co-workers (to Chisso); **Ger. Pat. 2,842,149** (Apr. 10, 1980), W. Pressler and H. Meidert (to Hoechst).
101. A. Hantzsch, *Justus Liebig Ann. Chem.* **215**, 1 (1882).
102. **S. Afr. Pat. 6801,482** and **U.S. Pat. 3,485,847** (Aug. 7, 1968) F. Bossert and W. Vater (to Bayer).
103. R. J. Clemens, *Kodak Lab. Chem. Bull.* **55**(3), (1984).
104. M. F. Carrol and A. R. Bader, *J. Am. Chem. Soc.* **75**, 460 (1953).
105. A. I. Meyers, W. R. Leonard, and J. L. Romine, *Tetrahedron Lett.* **32**, 597 (1991).
106. **Ger. Pat. 2,512,092** (Sept. 25, 1975), V. Hach (to Delmar Chemicals).
107. **Rus. Pat. 80,108,856** (Aug. 21, 1980), (to UBE Industries Ltd).
108. J. Perronet, P. Girault, and J. P. Demoute, *J. Heterocycl. Chem.* **17**, 727 (1980).
109. G. Falsone and B. Hundt, *Arch. Pharm.* **318**, 190 (1985); R. N. Lacey, *J. Chem. Soc.*, 822, 850 (1954).
110. J. L. Bloomer and F. E. Kappler, *J. Org. Chem.* **39**, 113 (1974); R. N. Lacey, *J. Chem. Soc.*, 832 (1954).
111. **Eur. Pat. Appl. 28,709** and **U.S. Pat. 4,473,508** (May 20, 1981), M. Gross (to Lonza AG).
112. A. Huwiler and C. Abächerli, *4-Haloacetoacetic Acid Derivatives*, Lonza AG Technical Brochure, Basel, Switzerland, 1990.
113. T. Ohta, T. Miyake, and H. Takaya, *J. Chem. Soc., Chem. Commun.*, 1725 (1992).
114. C. L. Perrin and T. Arrhenius, *J. Am. Chem. Soc.* **100**, 5249 (1978).
115. **Swiss Pat. Appl. 0157/91** (Jan. 21, 1991), C. O'Murchu and E. Greth (to Lonza AG).
116. T. Fujisawa, T. Sato, and Y. Gotoh, *Bull. Chem. Soc. Jpn.* **55**, 3555 (1982).
117. K. Itoh, T. Harada, and H. Nagashima, *Bull. Chem. Soc. Jpn.* **64**, 3746 (1991).
118. Ref. 6, p. 305.
119. R. Oda, S. Munemiya, and M. Okano, *Makromol. Chem.* **43**, 149 (1961).
120. E. V. Dehmlov and A. R. Shamout, *Justus Liebigs Ann. Chem.*, 2062 (1982).
121. **U.S. Pat. 2,585,537** (Feb. 12, 1952), D. D. Coffmann (to Du Pont).
122. Ref. 1, pp. 259, 276.
123. **U.S. Pat. 2,848,496** (Aug. 19, 1958), R. N. Lacey (to Distillers Co.).
124. **U.S. Pat. 3,865,846** (Feb. 11, 1975), G. Schulz and co-workers, (to BASF).
125. **Swiss Pat. 423,754** (May 13, 1967), H. Keller (to Lonza AG).
126. **Brit. Pat. 852,865** (Nov. 2, 1960), W. Moser and R. Perren (to Lonza AG).
127. **U.S. Pat. 3,759,955** (Sept. 18, 1973), M. L. Jacobs and B. W. Higdon (to Celanese).
128. **Brit. Pat. 3,271,420** (Dec. 31, 1965), H. Zima (to Lonza AG).
129. U.S. Department of Transportation classification: flammable liquid, hazard label 3.3; DOT UN Na shipping number 2521, flammable liquid. Transport by sea: IMDG-code class 3.3, p. 3333, emergency sheet 3–06. Transport by rail and road: RID ADR class 3 Rn 301-31c.
130. J. S. Witzeman, *Tetr. Letters* **31**, 1401 (1990).
131. Ref. 1, p. 229; G. Hommel, *Handbuch der gefährlichen Güter*, 5th ed., Springer, Berlin, 1993, Sheet Nr.350.
132. B. L. Van Duuren and co-workers, *J. Natl. Cancer Inst.* **31**, 41 (1963).
133. *Ibid.*, **37**, 825 (1966).
134. *Ibid.*, **39**, 1213, 1217 (1967).
135. Hoechst internal safety data sheet, Frankfurt, Germany, Oct. 1978.
136. **U.S. Pat. 3,733,413** (May 15, 1973), D. R. Spiner and R. K. Hoffmann (to U.S. Army).
137. **Fr. Pat. 1,579,873** (Aug. 29, 1969), A. C. Pierce (to Allied).

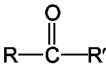
138. **Brit. Pat. 927,637** (May 29, 1963) (to Esso).
139. **Rus. Pat. 61,215,796** (Sept. 25, 1986), K. Mogi and co-workers (to Tomoegawa Paper Mfg.).
140. **Eur. Pat. A-442-431** (Aug. 21, 1991) and **Eur. Pat. A 444-563** (Sept. 4, 1991), T. Scholl and B. Jackson (to Lonza AG).
141. Ref. 2, p. 1485.
142. J. Fabian, H. Nakazumi, and H. Matsuoka, *Chem. Rev.* **92**, 1197 (1992).
143. G. Seitz and P. Imming, *Chem. Rev.* **92**, 1227 (1992).
144. Lonza AG, *Acetoacetates*, technical brochure, Basel, Switzerland, 1987.
145. *Beilstein Handbuch der Organischen Chemie*, 4th ed., Vol. 3, 1962, pp. 1178–1228.
146. *Ullmann's Encyklopädie der Technischen Chemie*, 4th ed. Vol. 13, 1977, pp. 157–159.
147. E. Dane and co-workers, *Angew. Chem.* **74**, 873 (1962); E. Dane and T. Dockner, *Angew. Chem.* **76**, 342 (1964); *Int. Ed.* **3**, 439 (1964).
148. R. Gasser, *Z. Naturf.* **8b**, 225 (1953).
149. **U.S. Pat. 2,754,243** (July 10, 1956), H. Gysin and A. Margot (to J. R. Geigy AG).
150. H. C. Bucha and co-workers, *Science* NY **137**, 537 (1962).
151. **Belg. Pat. 625,897** (Jun. 10, 1963), H. M. Loux, R. W. Luckenbaugh, and E. J. Soboczinski (to E. I. DuPont de Nemours & Co., Inc.); **U.S. Pat. 3,352,862** (Nov. 14, 1967), E. J. Soboczinski (to E. I. Du Pont de Nemours & Co., Inc.).
152. Technical brochure publication No. N-319, Eastman Kodak Co., Kingsport, Tenn., Oct. 1988; Lonza AG Product information sheet P5305, Lonza AG, Basel, Switzerland, 1993.
153. **U.S. Pats. 2,802,855** (Aug. 13, 1957), R. Whetstone and A. R. Stiles (to Shell), **3,400,177** (Sept. 3, 1968), L. F. Ward and J. G. Morales (to Shell).
154. R. Anliker and co-workers, *Helv. Chim. Acta* **44**, 1622 (1961); **U.S. Pat. 2,908,605** (Oct. 13, 1959), E. Beriger and R. Sallmann (to Ciba).
155. **Fr. Pat. 1,594,628** and **U.S. Pat. 3,908,008** (July 17, 1970) (to Bayer).
156. **U.S. Pat. 4,931,514** (Jun. 05, 1990) D. Waters.
157. **U.S. Pat. 3,249,499** (May 3, 1966), B. von Schmeling and co-workers (to Uniroyal).

Claudio Abaecherli
Lonza AG

Raimund J. Miller
Lonza, Inc.

KETONES

Ketones are a class of organic compounds that contain one or more carbonyl groups bound to two aliphatic, aromatic, or alicyclic substituents, and are represented by the general formula



Ketones are named by selecting as the parent compound the longest carbon chain that contains the carbonyl group, and by replacing the terminal "-e" of the parent compound by "-one". The parent chain is numbered in the direction which gives the carbonyl group the lowest number. CH₃COCH₂CH₃ thus becomes 2-butanone. In naming complex carbonyl structures containing more than one functional group, the carbonyl group takes precedence over alkene, hydroxyl, and most other groups. An older system of naming ketones simply lists the names of the R and R' groups attached to the carbonyl group, followed by the word "ketone". 2-Butanone is named methyl ethyl ketone using this methodology.

Ketones are an important class of industrial chemicals that have found widespread use as solvents and chemical intermediates. Acetone (qv) is the simplest and most important ketone and finds ubiquitous use as a solvent. Higher members of the aliphatic methyl ketone series (eg, methyl ethyl ketone, methyl isobutyl ketone, and methyl amyl ketone) are also industrially significant solvents. Cyclohexanone is the most important cyclic ketone and is primarily used in the manufacture of γ-caprolactam for nylon-6 (see CYCLOHEXANOL AND CYCLOHEXANONE). Other ketones find application in fields as diverse as fragrance formulation and metals extraction. Although the industrially important ketones are reviewed herein, the laboratory preparation of ketones is covered elsewhere (1).

Physical Properties

The lower molecular weight aliphatic ketones and cycloaliphatic ketones are stable, colorless liquids and generally have a pleasant, slightly aromatic odor. They are relatively volatile with boiling points slightly above those of corresponding alkanes. Unsymmetrical ketones are lower melting and higher boiling than corresponding symmetrical ketones. The members of the series up to C₅ are fairly soluble in water and are excellent solvents for nitrocellulose, vinyl resin lacquers, cellulose ethers and esters, and various natural and synthetic gums and resins.

In contrast, aromatic ketones are high boiling, colorless liquids that generally have a fragrant odor and are almost insoluble in water. They are useful as intermediates in chemical manufacture. Functionalized and cyclic ketones are also good solvents. Ring size and the type and location of functional groups affect odor, color, and reactivity of these ketones.

The physical properties of some common ketones are listed in Table 1. Ketones are commonly separated by fractional distillation, and vapor–liquid equilibria and vapor pressure data are readily available for common ketones. A number of other temperature dependent physical properties for acetone, methyl ethyl ketone, methyl isobutyl ketone, and diethyl ketone have been published (3).

Table 1. Physical Properties of Ketones^a

Systematic name (trivial or common name)	CAS Registry Number	Mol wt	Freezing point, °C	Boiling point at 101.3 kPa, ^b °C	Refractive index, <i>n</i> _D ²⁰	Specific gravity 20/20, °C	Viscosity at 20°C, mPas (cP)	Surface tension at 20°C	Heat of vaporization at 101.3 kPa, ^b kJ mol ^c	Liquid specific heat capacity at (T)°C, J (kgK) ^c	Flash point, open cup, (closed)	Solubility at 20°C, wt %	
												In water	Water in
<i>Methyl alkyl ketones</i>													
2-propanone (acetone)	[67-64-1]	58.08	94.7	56.1	1.3590	0.7905	0.33	24.0	29.53	2224 (30)	16 (18)	complete	complete
2-butanone (methyl ethyl ketone)	[78-93-3]	72.10	85.9	79.57	1.3780	0.8062	0.41	24.6	31.64	2203 (20)	6 (6)	26.8	11.8
2-pentanone (methyl propyl ketone)	[107-87-9] /	86.13	77.8	102.4	1.3902	0.8076	0.51 (28.3°C)	23.2	33.39		(7)	4.3	3.3
3-methyl-2-butanone (methyl isopropyl ketone)	[563-80-4] /	86.13	92	94.2	1.3882	0.8044	0.43 (25°C)	24.6 (25°C)	30.63		(6)	6.53	
4-methyl-2-pentanone (methyl isobutyl ketone)	[108-10-1] /	100.16	84.0	116.2	1.3957	0.8020	0.61	23.6	35.60	1920 (20)	23 (16)	1.6	1.9
2-hexanone (methyl <i>n</i> -butyl ketone)	[591-78-6] /	100.16	55.8	127.5	1.4007	0.8125	0.62	25.4	36.05	2228 (25)	(35)	1.75	3.7
3-methyl-2-pentanone (methyl <i>sec</i> -butyl ketone)	[565-61-7] /	100.16	83	117.4	1.4001	0.8142			35.12			2.26	
3,3-dimethyl-2-butanone (pinacolone)	[75-97-8] /	100.16	50	106.4	1.3986	0.8070			33.5			2.0	1.8
2-heptanone (methyl amyl ketone)	[110-43-0] /	114.18	35	151.5	1.4087	0.8166	0.77	26.1	39.25		47 (49)	0.43	1.45
5-methyl-2-hexanone (methyl isoamyl ketone)	[110-12-3] /	114.18	73.9	144.9	1.4069	0.8127	0.77 (25°C)	25.3 (25°C)			41 (35)	0.54	1.28
2-octanone (methyl hexyl ketone)	[111-13-7] /	128.22	20.5	173.3	1.4153	0.8197	0.95 (25°C)	26.6 (25.5)	40.88		(62)		

4-hydroxy-4-methyl-2-pentano ne (diacetone alcohol)	[123-42-2 /]	116.1 6	44.2	169.2	1.422 6	0.9406	3.2	31	41.6	1883	61 (47)	compl ete	compl ete
<i>Dialkyl ketones</i>													
3-pentanone (diethyl ketone)	[96-22-0/]	86.3	39.4	101.8	1.392 3	0.8155	0.47	24.7	33.69	2215 (25)	(13)	3.4	2.6
2,4-dimethyl-3-pentanone (diisopropyl ketone)	[565-80-0 /]	114.1 9	69	125.0	1.399								
2,6-dimethyl-4-heptanone (diisobutyl ketone)	[108-83-8 /]	142.2 4	46	169.4	1.417 2	0.8076	1.02	22.2	39.31		49 (49)	0.05	0.75
3-hexanone (ethyl propyl ketone)	[589-38-8 /]	100.1 6		123.2	1.400 3	0.8174		25.04	35.66		35	1.57	
3-heptanone (butyl ethyl ketone)	[106-35-4 /]	114.1 9	39	147.3	1.408 8	0.8197	0.84	25.7	36.59		41 (46)	0.43	0.78
3-octanone (ethyl amyl ketone)	[106-68-3 /]	128.2 2	46	167–1	1.415 0	0.8220							
2,6,8-trimethyl-4-nonanone (isobutyl heptyl ketone)	[123-18-2 /]	184.3 2	75	218.2	1.425 7	0.8180	1.9		44.56		90 (88)	<0.01	0.2
<i>Unsaturated ketones</i>													
3-buten-2-one (methyl vinyl ketone)	[78-94-4/]	70.09	6	81.4	1.413 0								
3-methyl-2-buten-2-one (methyl isopropenyl ketone)	[814-78-8 /]	84.12	54	98	1.423 6	0.855							
4-methyl-3-penten-2-one (mesityl oxide)	[141-79-7 /]	98.15	53	129.5	1.441 4	0.8521	0.6	28.4	43.1	2176 (20)	29 (31)	3.1	3.4
4-methyl-4-penten-2-one (isomesityl oxide)	[3744-02- 3/]	98.15		121.5	1.445 8	0.8548							
3,5,5-trimethyl-2-cyclo-hexen-1 -one (isophorone)	[78-59-1/]	138.2 1	8.1	215.3	1.477 5	0.9229	2.6	32	43.4	1799 (20)	104 (85)	4.3(25 °C)	1.2(25 °C)
3,5,5 trimethyl-3-cyclo-hexen-1-one (β-isophorone)	[471-01-2 /]	138.2 1		181–1		0.89						0.03	
<i>Diketones</i>													
2,3-butanedione (diacetyl)	[431-03-8 /]	86.09	2.5	90.2	1.393 8	0.9843			34.3				
2,3-pentanedione	[600-14-6 /]	100.1 2	52	111				31	35.4	1983 (20)			
2,4-pentanedione (acetylacetone)	[123-54-6 /]	100.1 1	23.5	140.4	1.451 0	0.9753	0.58		36.55	1956.2			
2,5-hexandione	[110-13-4 /]	114.1 5	5.4	192.3	1.425 6	0.9734	1.6			(15)			
<i>Cyclic ketones</i>													
cyclopentanone (adipic ketone)	[120-92-3 /]	84.12	50.6	130.8	1.435 9	0.9512	1.2	33.35	36.53			29	14
cyclohexanone (pimelic ketone)	[108-94-1 /]	98.15	31.1	155.7	1.451 0	0.9482	2.21	35.2	37.62	2039.8	46 (43)	2.5	8.0
										(30.8°C)			
cycloheptanone	[502-42-1 /]	112.1 7	21	179	1.461 1			26.4			72 (62.5)	0.3	1.4
3,3,5-trimethylcyclohexanone	[873-94-9 /]	140.2 2	10	188.8	1.445 5	0.888	2.54						
<i>Aromatic ketones</i>													
acetophenone (methyl phenyl ketone)	[98-86-2/]	120.1 5	19–20	201.7	1.534 2	1.0296	0.93		45.69		93 (82)	0.55	1.65
benzophenone (diphenyl ketone)	[119-61-9 /]	182.2 2	48–49	305									
1-phenyl-2-propanone (phenylacetone)	[103-79-7 /]	134.1 8	15		1.515 8								
propiophenone (phenyl ethyl ketone)	[93-55-0/]	134.1 7	18.2	218	1.526 5	1.012		37.4	45.44		96 (85)	0.01(2 5°C)	

^a Ref. 2.
^b To convert kPa to mm Hg, multiply by 7.5.
^c To convert J to cal, divide by 4.184.

Ketones, like aldehydes, tend to form azeotropes with water and other substances. A table of ketone–water azeotropes for some commercially produced ketones is listed in Table 2.

Table 2. Ketone–Water Azeotropes ^a

Ketone	Ketone bp at 101.3 kPa, ^b °C	Azeotrope bp at 101.3 kPa, ^b °C	Azeotrope ketone composition, wt %
acetone	84 ^c	81.4	98.7
acetophenone	201.7	99.1	18.5
cyclohexanone	155.4	95	38.4
diacetone alcohol	169.2	99.6	13
diisobutyl ketone	169.4	97	48
isobutyl heptyl ketone	218.2	99	16

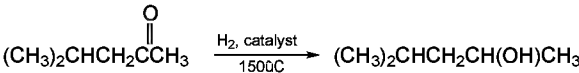
isophorone	215.2	99.5	16.1
mesityl oxide	129.8	91.8	65.3
methyl ethyl ketone	79.6	73.4	88
methyl isobutyl ketone	116.2	87.9	76
2,4-pentanedione	140.4	94.4	59

^a Ref. 2.
^b To convert kPa to mm Hg, multiply by 7.5.
^c At 137.9 kPa.

Chemical Properties

The constituent carbonyl group makes many of the reactions and methods of preparation for ketones similar to those of aldehydes. Ketones, however, generally undergo 1,2-addition reactions across the carbonyl group less readily than aldehydes because of steric hindrance around the carbonyl group. Similarly, the relative reactivity among ketones is influenced by the polarity and electrophilic nature of the substituents in the vicinity of the carbonyl group (eg, hydrogens alpha to the carbonyl group). The chemical properties of diketones, and cyclic and unsaturated ketones such as 2,4-pentanedione, cyclohexanone, and mesityl oxide, respectively, are enhanced, thereby increasing their utility as chemical intermediates.

Reduction. Most ketones are readily reduced to the corresponding secondary alcohol by a variety of hydrogenation processes. The most commonly used catalysts are palladium, platinum, and nickel. For example, 4-methyl-2-pentanol (methyl isobutyl carbinol) is commercially produced by the catalytic reduction of 4-methyl-2-pentanone (methyl isobutyl ketone) over nickel.



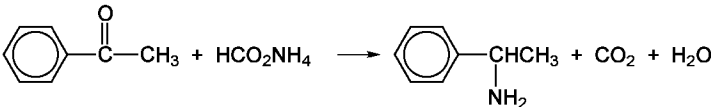
Oxidation. Ketones are oxidized with powerful oxidizing agents such as chromic or nitric acid. During oxidation, carbon–carbon bond cleavage occurs to produce carboxylic acids. Ketone oxidation with hydrogen peroxide, or prolonged exposure to air and heat, can produce peroxides. Concentrated solutions of ketone peroxides (>30%) may explode, but dilute solutions are useful in curing unsaturated polyester resin mixtures (see PEROXIDES AND PEROXY COMPOUNDS).

Condensation.

Base Catalyzed. Depending on the nature of the hydrocarbon groups attached to the carbonyl, ketones can either undergo self-condensation, or condense with other activated reagents, in the presence of base. Name reactions which describe these conditions include the aldol reaction, the Darzens-Claisen condensation, the Claisen-Schmidt condensation, and the Michael reaction.

Acid Catalyzed. Although ketonic carbonyl groups are less reactive than aldehydic carbonyls in the presence of basic catalysts, this is not the case with acid catalysts. Thus acetone undergoes aldol addition in the presence of sulfuric acid to give mesityl oxide, which then condenses with a third molecule of acetone to give a mixture of phorone (2,6-dimethyl-2,6-heptadien-4-one) and mesitylene (1,3,5-trimethylbenzene). Ketones also condense with activated aromatic products in the presence of sulfuric acid to give coupled aromatic products. For example, acetone and phenol condense to bisphenol A (4,4'-isopropylidenediphenol), which is used in the manufacture of epoxy resins (qv) and polycarbonates (qv).

Preparation of Amines. Amines are prepared by heating aliphatic, aromatic, or cyclic ketones with ammonium formate, formamide, or an N-substituted ammonium formate at 165–190°C (Leuckart reaction). For example, α-methylbenzylamine is prepared by the reaction of acetophenone with ammonium formate.



diacetone alcohol	4,000		13.5		5		50
methyl- <i>n</i> -amyl ketone	1,670		14 mg 24 h MLD	4,000			100
diisobutyl ketone	5,750	16		2,000			25, 50 (trans)
ethyl butyl ketone	2,760		500 mg 24 h MOD	2,000	100 ^h		50
2,4-pentane-dione	1,000	5		1,000		4.760	
mesityl oxide	1,120		430 mg open MLD	ⁱ		4.325	15, 25 (trans)
isophorone	2,330		1.5	1,840		0.920	4, 25 (trans)
cyclohexanone	1,535		948 mg kg	8,000 ^f		4.740	25, 50 (trans)
acetophenone	815		515 mg open MLD			0.771	

^a Ref. 5.

^b MLD mild, MOD moderate.

^c The lowest concentration in air that caused death in four hours; LC_{L0} unless otherwise noted.

^d Severe unless otherwise noted.

^e 8-h time-weighted average; trans transitional.

^f LC₅₀.

^g mg m⁻³.

^h Mild.

ⁱ LC₅₀ 9 g m⁻³ 4 h.

Environmental Aspects

Most industrially important ketones are volatile organic compounds (VOCs) which are subject to air pollution control regulations. The purpose of these regulations is to limit the atmospheric release of materials which could either be toxic or could be precursors of ozone and other species associated with photochemical smog. Legislation has become progressively more stringent as the long-term effects of atmospheric pollution have become evident. In the United States, most states have developed their own programs to govern the release of toxic air pollutants (TAPs). Pertinent federal regulations include the 1986 Superfund Amendments and Reauthorization Act (SARA): Title III (Section 313); the 1990 Clean Air Act Amendments; and the voluntary agreement between the Environmental Protection Agency (EPA) and the Chemical Manufacturers Association (CMA) termed the 33-50 project (see AIR POLLUTION).

The 1990 Clean Air Act Amendments list 189 hazardous air pollutants (HAPs) that the EPA must regulate to enforce maximum achievable control technology (MACT) to standards which are to be set by the year 2000. The 33-50 project calls for reduction of emissions of 17 specified solvents to predetermined levels by 1995. The SARA statute provides a mechanism by which the community can be informed of the existence, quantities, and releases of toxic chemicals, and requires that anyone releasing specific toxic chemicals above a threshold level to annually submit a toxic chemical release form to the EPA. The status of various ketones under these regulations is shown in Table 4.

Table 4. Air Pollution Regulations Affecting Ketones

Ketone	VOC ^a	HAP	33-50 ^b	SARA 313 ^c
acetone	yes			yes
methyl ethyl ketone	yes	yes	yes	yes
methyl isobutyl ketone	yes	yes	yes	yes
diacetone alcohol	yes			
diisobutyl ketone	yes			
methyl amyl ketone	yes			
isophorone	yes	yes		
cyclohexanone	yes			
acetophenone	yes	yes		

^a Ref. 7.

^b Ref. 8.

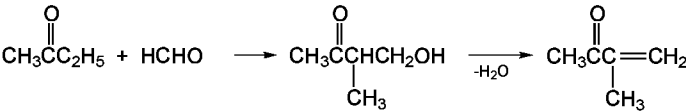
^c Ref. 9.

The impact of the regulations in Table 4 is to require users and producers of VOC ketones to limit release by either reformulating to new solvent systems, to install environmental control systems which recover and recycle solvents, or reduce emissions with carbon absorption beds or incineration equipment. The use of some individual ketones will decline further, but the overall short-term use of ketones is forecast to remain stable (10).

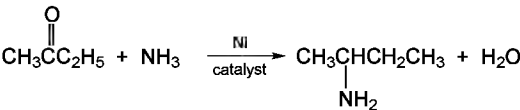
Aliphatic Ketones

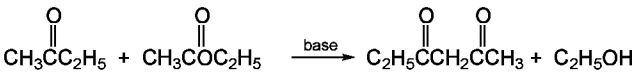
Methyl Ethyl Ketone. Methyl ethyl ketone (MEK) (2-butanone), CH₃COCH₂CH₃, is the next higher aliphatic ketone homologue to acetone, and third to acetone and cyclohexanone as the most important commercially produced ketone.

MEK is a colorless, stable, flammable liquid possessing the characteristic acetone-type odor of low molecular weight aliphatic ketones. MEK undergoes typical reactions of carbonyl groups with activated hydrogen atoms on adjacent carbon atoms, and condenses with a variety of reagents. Condensation of MEK with formaldehyde produces methylisopropenyl ketone (3-methyl-3-buten-2-one):



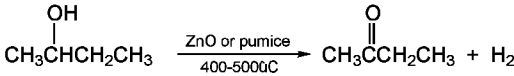
Reaction of MEK with ammonia and hydrogen produces *sec*-butylamine, a fungistat and condensation with aliphatic esters under strongly alkaline conditions produces 1,3-diketones.





Direct oxidation yields biacetyl (2,3-butanedione), a flavorant, or methyl ethyl ketone peroxide, an initiator used in polyester production.

Manufacture. MEK is predominantly produced by the dehydrogenation of 2-butanol. The reaction mechanism (11–13) and reaction equilibrium (14) have been reported, and the process is in many ways analogous to the production of acetone (qv) from isopropyl alcohol.



The 2-butanol feedstock is conventionally obtained by the sulfuric acid-catalyzed addition of water to *n*-butenes. This is a two-step reaction involving sulfation and hydrolysis in which the conversion of *n*-butenes is 90% and selectivity to 2-butanol is 95% (15). During operation the sulfuric acid becomes diluted and must be reconcentrated before reuse. In 1983 Deutsche Texaco commercialized a single-step route in which 2-butanol is formed by the hydration of *n*-butenes in the presence of a strongly acidic ion-exchange resin containing sulfonic acid groups (16–18). The direct reaction is carried out at 150–160°C and 7 MPa. Virtually anhydrous 2-butanol is recovered in this process (19). Direct hydration requires lower utilities and investment costs, operates at 99% selectivity to 2-butanol, but is hindered by low (5–15%) *n*-butene conversion per pass.

The dehydrogenation of 2-butanol is conducted in a multitube vapor-phase reactor over a zinc oxide (20–23), copper (24–27), or brass (28) catalyst, at temperatures of 250–400°C, and pressures slightly above atmospheric. The reaction is endothermic and heat is supplied from a heat-transfer fluid on the shell side of the reactor. A typical process flow sheet is shown in Figure 1 (29). Catalyst life is three to five years; operating in three to six month cycles between oxidative reactivations (30). Catalyst life is impaired by exposure to water, butene oligomers, and di-*sec*-butyl ether (27).

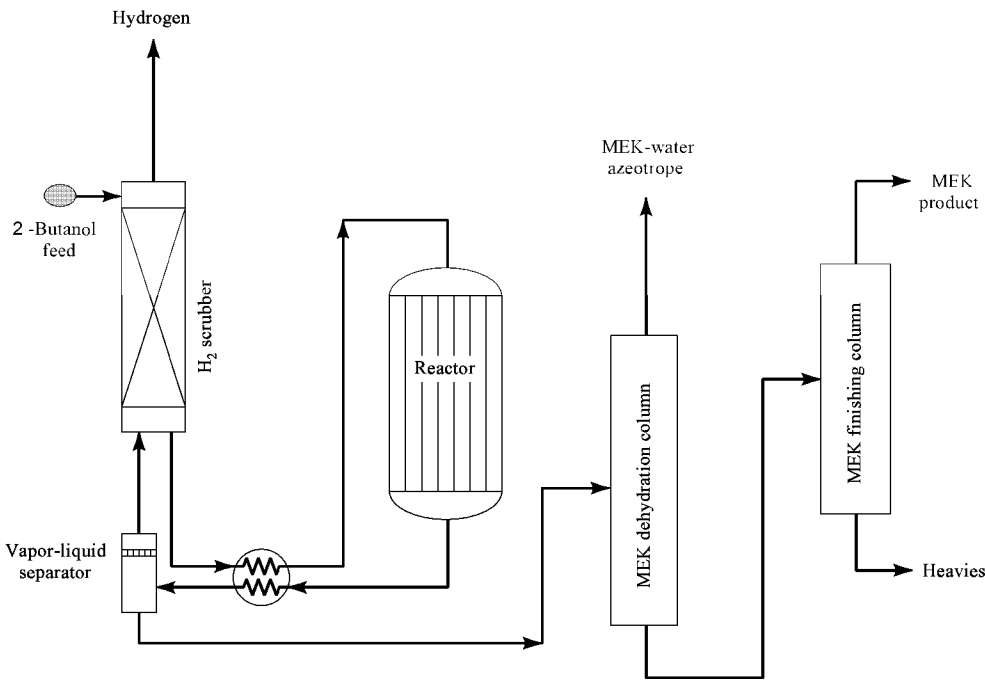


Fig. 1. Dehydrogenation of 2-butanol to methyl ethyl ketone (29).

Courtesy of SRI International.

MEK is also produced as a by-product in the liquid-phase oxidation of *n*-butane to acetic acid (31–33). This route was once the most favored route to acetic acid, however, since the early 1980s the acetic acid technology of choice has become methanol carbonylation, and MEK growth by this path is doubtful.

Another attractive commercial route to MEK is via direct oxidation of *n*-butenes (34–39) in a reaction analogous to the Wacker-Hoechst process for acetaldehyde production via ethylene oxidation. In the Wacker-Hoechst process the oxidation of olefins is conducted in an aqueous solution containing palladium and copper chlorides. However, unlike acetaldehyde production, *n*-butene oxidation has not proved commercially successful because chlorinated butanones and butyraldehyde by-products form which both reduce yields and complicate product purification, and also because titanium-lined equipment is required to withstand chloride corrosion.

A variation of the Pd–Cu Wacker-Hoechst process, termed OK Technology, has been proposed by Catalytica Associates (40–46). This process avoids the use of chlorides and uses a Pd–Cu catalyst system which incorporates a polyoxoanion and a nitrile ligand.

Economic Aspects. The 1992 MEK nameplate capacity for the United States, East Asia, and Western Europe is listed in Table 5. During the period 1980–1989 MEK achieved a negative growth rate as demand dropped from 311,000 (48) to 228,000 t/yr (49). Stricter VOC regulations were largely responsible for the decline, and the trend will continue as solvent recovery and recycling, as well as substitution away from MEK, take effect.

Table 5. United States, East Asia, and Western Europe Methyl Ethyl Ketone Production^a

Producer	Plant location	1992 Capacity, 10 ³ t/yr
Exxon	Baton Rouge, La.	104.3 ^b
Hoechst-Celanese ^c	Pampa, Tex.	36.3 ^b
Shell	Norco, La.	104.3 ^b
U.S. total		244.9
Idemitsu Kosan	Tokukama, Japan	40
Maruzen Petro.	Ichihara, Japan	90
Tonen Chemical	Kawasaki, Japan	70
Lee Chang Yung	Linyuan, Taiwan	20

Taiwan Synth. Pet.	Linyuan, Taiwan	40
<i>Eastern Asia total</i>		<i>260</i>
Atochem SA	La Chambre, France	50
RWE (Deutsche Tex)	Moers, Germany	60
Exxon	Fawley, U.K.	100
Shell	Ellesmere, U.K.	20
	Berre, France	45 ^d
	Pernis, Holland	60
<i>Western Europe total</i>		<i>335</i>

^a Ref. 47. Process is dehydrogenation of 2-butanol unless otherwise noted.
Courtesy of SRI International.

^b 1993 data.

^c Process is butane oxidation.

^d Scheduled for closure March 1993 (55).

Domestic capacity declined correspondingly during the 1980s as ARCO closed its 33,000 t yr 2-butanol dehydrogenation plant in 1991 (50), Exxon closed a 136,000 t yr plant at Bayway, New Jersey in 1988, and Union Carbide closed its Brownsville, Texas plant. Consumption was further squeezed in 1988 as a result of an outage at the Hoechst-Celanese (Pampa, Texas) chemical plant in November 1987 (51,52). The plant came back on-stream in early 1989 (53). In addition, in May 1988 an incident at Shell's Norco plant (54) limited production. These outages hastened many customers to switch over to other solvent substitutes. The price of MEK in October 1994 was about \$0.88 /kg. Supply fluctuations caused the price to range from 0.55 to 1.10 \$ /kg since the early 1980s.

Health and Safety Factors. MEK is slightly more toxic than acetone, but is not considered highly toxic, and nor does it exhibit cumulative toxicological properties. The OSHA time weighted average in air is 200 ppm; other measured toxicity values are shown in Table 3. Methyl ethyl ketone is highly flammable.

Uses. A market breakdown for MEK is shown in Table 6. The foremost use of MEK is as a coating solvent because it provides low viscosity solutions at high solids content without affecting film properties. It is a suitable solvent for practically all synthetic and natural resins commonly employed in lacquers, and is distinguished from acetone uses by its lower solubility in water and its higher boiling point. Widespread use in the automotive, appliance, furniture, and housing industries will slowly decline as emphasis moves to high solids, uv-curable, and waterborne coatings.

Table 6. U.S. Consumption of Methyl Ethyl Ketone, 10³ t ^a

Use	1985	1988
coating solvent	129	154
adhesives	31	22
magnetic tapes	16	9
printing inks	10	4
lube oil dewaxing	12	13
intermediates	10	11
other	30	26
Total	238	239

^a Ref. 56. Courtesy of SRI International.

MEK is also used in solvent-based adhesives, in printing ink formulations, as a solvent in magnetic tape manufacture, and is the most common solvent used in dewaxing lubricating oils. Of all these applications only an increasing consumption of magnetic tapes is likely to grow in methyl ethyl ketone use.

Other uses include use as a reaction and extraction solvent in pharmaceutical production; as an intermediate for the preparation of catalysts, antioxidants (qv), and perfumes; and as a feedstock in the production of methyl isopropenyl ketone, 2,3-butanedione, and methyl ethyl ketone peroxide. Concern has also arisen at the large volume of exported MEK which has been covertly diverted and used to process cocaine in Latin American countries (57).

Methyl Isobutyl Ketone. Methyl isobutyl ketone (MIBK) (4-methyl-2-pentanone), (CH₃)₂CHCH₂COCH₃, is an industrially important solvent which after methyl methacrylate and bisphenol A is the third largest tonnage product obtained from acetone.

MIBK is a flammable, water-white liquid that boils at 116°C. It is sparingly soluble in water, but is miscible with common organic solvents. It forms an azeotrope with water as shown in Table 2. Condensation of MIBK with another methyl ketone can produce ketones containing 9–15 carbons. For example, condensation with acetone produces diisobutyl ketone, and self-condensation of two MIBK molecules produces 2,6,8-trimethyl-4-nonanone [123-17-1]. Condensation with 2-ethylhexanal gives 1-tetradecanol (7-ethyl-2-methyl-4-undecanol), a valuable surfactant intermediate (58).

The carbonyl functionality of MIBK can be hydrogenated over nickel catalysts to yield methyl isobutyl carbinol (4-methyl-2-pentanol or methyl amyl alcohol) [108-11-2]. Industrial processes coproduce methyl isobutyl carbinol during the hydrogenation of mesityl oxide to MIBK. The product ratio of methyl isobutyl carbinol–MIBK during this reaction can be shifted toward methyl isobutyl carbinol by adopting a higher than normal pressure and H₂:organic ratio (59). Methyl isobutyl carbinol is used as an ore flotation frother and to produce zinc dialkyl dithiophosphate lube oil additives. It is produced in the United States by Shell and Union Carbide (\$1.12 /kg, October 1994).

Manufacture. MIBK is produced industrially by three different routes: (1) in a three-step process using acetone and hydrogen as feeds and proceeding via diacetone alcohol and mesityl oxide intermediates; (2) in a one-step synthesis from acetone and hydrogen; and (3) in a one-step mixed ketone process from isopropyl alcohol. The three-step process was the conventional technology up until the late 1960s; thereafter the simpler single-step synthesis directly from acetone was commercialized by Veba-Chemie and Deutsche Texaco in Germany, and Tokuyama Soda in Japan. In addition, Eastman developed a proprietary unpatented one-step process from acetone which coproduces other by-products (60). Despite the improvements offered by the direct route many of the older three-step plants are still in operation.

The one-step route from 2-propanol coproduces diisobutyl ketone and acetone, and is practiced in the United States by Union Carbide (61). The details of a vapor-phase 2-propanol dehydrogenation and condensation process for the production of acetone, MIBK, and higher ketones have been described in recent patents (62,63). The process converts an azeotropic 2-propanol–water feed over a copper-based catalyst at 220°C and produces a product mixture containing 2-propanol (11.4° o), acetone (52.4° o), MIBK (21.6° o), diisobutyl ketone (6.5° o), and 4-methyl-2-pentanol (2.2° o).

In the three-step process acetone first undergoes a liquid-phase alkali-catalyzed condensation to form diacetone alcohol. Many alkali metal oxides, metal hydroxides (eg, sodium, barium, potassium, magnesium, and lanthanum), and anion-exchange resins are described in the literature as suitable catalysts. The selectivity to diacetone alcohol is typically 90–95 wt % (64). In the second step diacetone alcohol is dehydrated to mesityl oxide over an acid catalyst such as phosphoric or sulfuric acid. The reaction takes place at 95–130°C and selectivity to mesityl oxide is 80–85 wt % (64). A one-step conversion of acetone to mesityl oxide is also possible.

Finally, selective hydrogenation of the olefinic bond in mesityl oxide is conducted over a fixed-bed catalyst in either the liquid or vapor phase. In the liquid phase the reaction takes place at 150°C and 0.69 MPa, in the vapor phase the reaction can be conducted at atmospheric pressure and temperatures of 150–170°C. The reaction is highly exothermic and yields 8.37 kJ mol (65). To prevent temperature runaways and obtain high selectivity, the conversion per pass is limited in the liquid phase, and in the vapor phase inert gases often are used to dilute the reactants. The catalysts employed in both vapor- and liquid-phase processes include nickel (66–76), palladium (77–79), copper (80,81), and rhodium hydride complexes (82). Complete conversion of mesityl oxide can be obtained at selectivities of 95–98%.

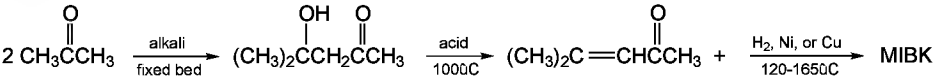


Figure 2 illustrates the three-step MIBK process employed by Hibernia Scholven (83). This process is designed to permit the intermediate recovery of refined diacetone alcohol and mesityl oxide. In the first step acetone and dilute sodium hydroxide are fed continuously to a reactor at low temperature and with a reactor residence time of approximately one hour. The product is then stabilized with phosphoric acid and stripped of unreacted acetone to yield a crude diacetone alcohol stream. More phosphoric acid is then added, and the diacetone alcohol dehydrated to mesityl oxide in a distillation column. Mesityl oxide is recovered overhead in this column and fed to a further distillation column where residual acetone is removed and recycled to yield a tails stream containing 98–99% mesityl oxide. The mesityl oxide is then hydrogenated to MIBK in a reactive distillation conducted at atmospheric pressure and 110°C. Simultaneous hydrogenation and rectification are achieved in a column fitted with a palladium catalyst bed, and yields of mesityl oxide to MIBK exceeding 96% are obtained.

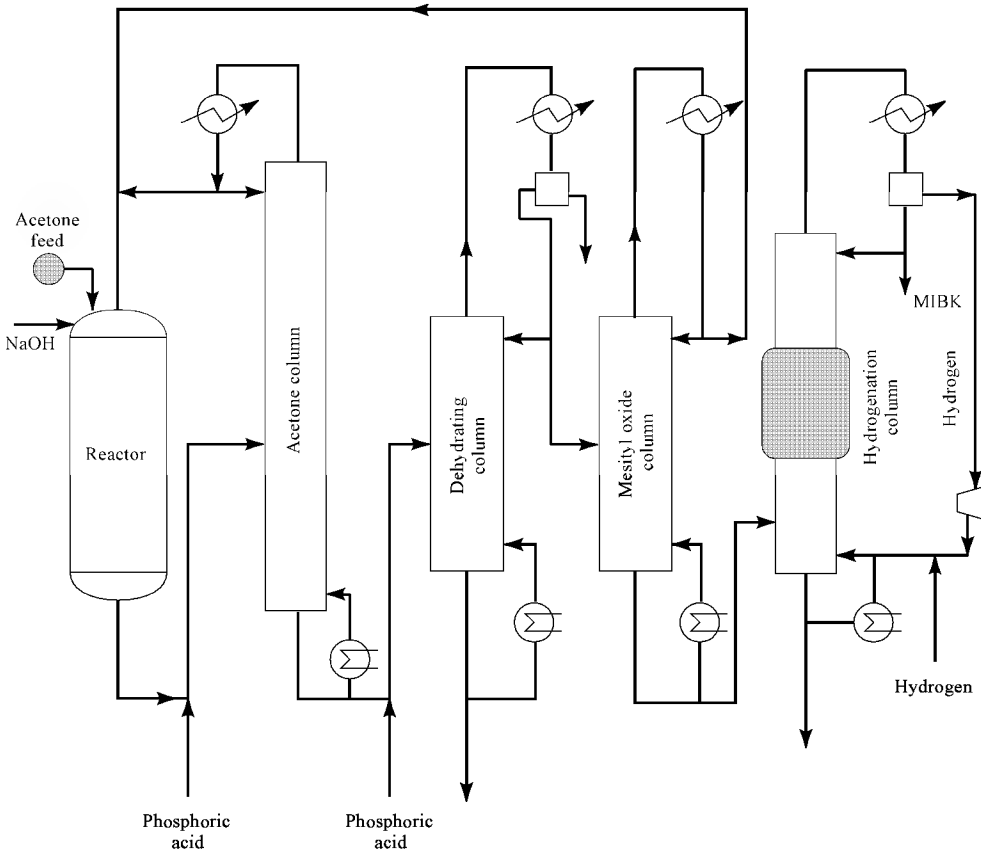


Fig. 2. Three-step MIBK process (83).
Courtesy of Chemische Industrie.

The single-step manufacture of MIBK offers lower investment and operating costs, and avoids the low conversion of acetone in the first stage and the reversion of mesityl oxide to acetone in the second stage, which are experienced in the three-step process. Direct synthesis is achieved using a multifunctional catalyst which effects acid aldolization, dehydration, and hydrogenation. Veba-Chemie's patents (84–86) describe a catalyst composed of a cation-exchange resin loaded with 0.05% palladium, and over which is passed a 2:1 mole feed ratio of H₂:acetone at approximately 135°C and 6.2 MPa. At these conditions a 96% selectivity to MIBK is achieved at 35% acetone conversion. Refined MIBK is then recovered from a four-column refining train in which the first column removes light hydrocarbons, and the second recycles unconverted acetone. A decanter is then located upstream of the final two columns and is used to separate an aqueous phase. The third column removes 2-propanol–water mixture, and the final column produces refined MIBK overhead and a heavies (diisobutyl ketone) tails stream. A similar process is operated by Deutsche Texaco (87–90) at operating conditions of 130–140°C and 3 MPa. An acetone conversion of 40% and a consumption of 1.4 kg of acetone per kg of MIBK is reported (91). A flow sheet of the Deutsche Texaco process is shown in Figure 3 (92).

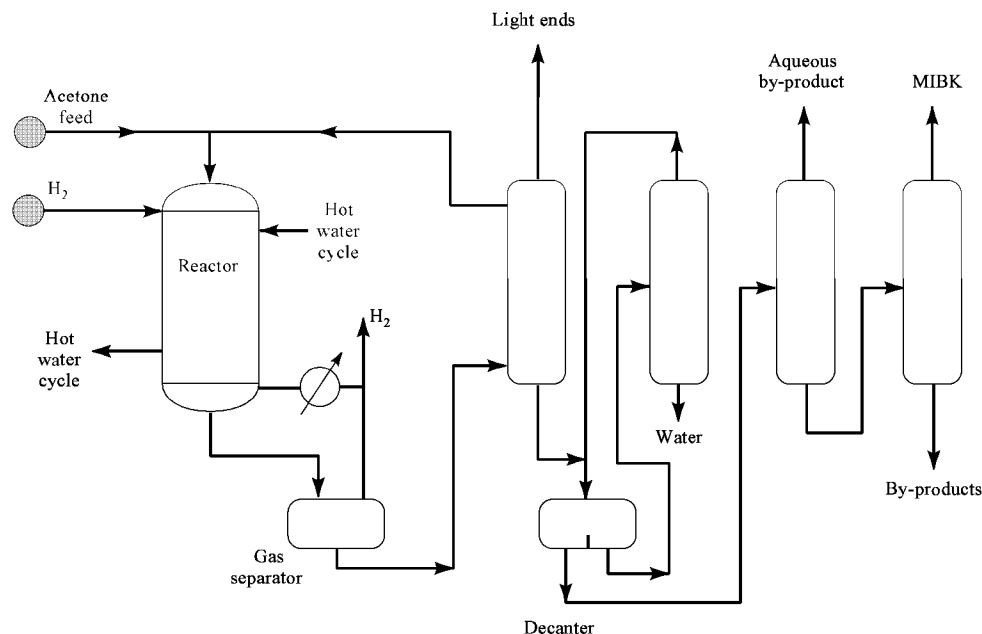


Fig. 3. Deutsche Texaco one-step MIBK process (92).

The Tokuyama Soda single-step catalyst consists of a zirconium phosphate catalyst loaded with 0.1–0.5 wt % palladium (93–97). Pilot-plant data report (93) that at 140°C, 3 MPa, and a H₂:acetone mole ratio of 0.2, the MIBK selectivity is 95% at an acetone conversion of 30%. The reactor product does not contain light methyl substituted methyl pentanes, and allows MIBK recovery in a three-column train with a phase separator between the first and second columns.

Sumitomo Chemical Co. (98–100) and Mitsubishi Kasei Co. (101) have patented single-step catalysts containing niobium and palladium. A Sumitomo example reports 93.5% MIBK selectivity at 41.8% acetone conversion and conditions of 160°C and 2 MPa. Other significant processes have been reported (60,102–110).

Economic Factors. The United States, Western European, and Japanese MIBK production capacities are shown in Table 7. During 1980–1989, United States domestic capacity remained relatively constant as suppliers offset the handicap of MIBK's status as a nonexempt VOC solvent through market gains in increasingly popular high solids coatings formulations. Growth during this period was 0.2% per year, and a small growth of 2% per year was forecast over the period 1990–1994 (111). Consumption of MIBK experienced a short-term strengthening in 1987 and 1988 as formulations of MIBK and acetone found use as a replacement for methyl ethyl ketone, which was experiencing acute supply shortages due to plant outages. MIBK was priced in the range \$1–1.12/kg as of October 1994.

Table 7. United States, East Asia, and Western Europe Methyl Isobutyl Ketone Production^a

Producer	Plant location	1992 Capacity, 10 ³ t/yr
Eastman	Kingsport, Tenn.	18 ^b
Shell	Deer Park, Tex.	45 ^b
Union Carbide	Institute, W.V.	34 ^b
<i>U.S. total</i>		97
Kyowa Yuka	Yokkaichi, Japan	15
Mitsubishi Kasei	Mizushima, Japan	17
Mitsui Petro.	Otake, Japan	24
Kumho Shell	Yeochan, South Korea	12
Lee Chang Yung	Linyuan, Taiwan	10
<i>Eastern Asia total</i>		78
Atochem	La Chambre, France	14
Höls	Herne, Germany	10
RWE (Deutsche Texaco)	Moers, Germany	8
Shell	Berre, France	25
	Pernis, Netherlands	25
	Stanlow, England	
<i>Western Europe total</i>		82

^a Ref. 47. Courtesy of SRI International.

^b 1993 data.

Health and Safety Factors. Like other low molecular weight ketones, MIBK is an anesthetic chemical with no highly cumulative toxicological effects. Inhalation of vapors can irritate mucous membranes.

Uses. The principal uses of MIBK are categorized in Table 8. Like methyl ethyl ketone, the principal use of MIBK is as a coating solvent. As a solvent for cellulose-based (eg, nitrocellulose and cellulose acetate butyrate) and resin-based (eg, acrylic, alkyd, and vinyl) coating systems, MIBK is unsurpassed. Attempts to replace MIBK with straight-chain solvents which are exempt from Rule 66 of the Los Angeles, California Air Pollution Control District, and other regulations restricting emission of photoreactive organic materials, have not been implemented as rapidly as expected. MIBK's solvent use has been sustained by the increasing use of high solids coatings, which use less solvent and require the superior solvent qualities offered by MIBK. Another increasingly important non-VOC use of MIBK is as a raw material in the production of rubber antioxidants such as N-(1,3-dimethylbutyl)-N'-phenyl-p-phenylenediamine.

Table 8. Uses of Methyl Isobutyl Ketone, 1988^a

Uses	10 ³ t
------	-------------------

surface applications	
surface coatings	45.4
solvent extraction	4.5
other	5.9
rubber antioxidants	11.3
specialty surfactants	1.4
<i>Total</i>	<i>68.5</i>

^a Ref. 61. Courtesy of SRI International.

MIBK is a highly effective separating agent for metals from solutions of their salts and is used in the mining industries to extract plutonium from uranium, niobium from tantalum, and zirconium from hafnium (112,113). MIBK is also used in the production of specialty surfactants for inks (qv), paints, and pesticide formulations, examples of which are 2,4,7,9-tetramethyl-5-decyn-4,7-diol and its ethoxylated adduct. Other applications include as a solvent for adhesives and wax oil separation (114), in leather (qv) finishing, textile coating, and as a denaturant for ethanol formulations.

Diacetone Alcohol. Diacetone alcohol (DAA) (4-hydroxy-4-methyl-2-pentanone) is a colorless, mild smelling liquid which is completely miscible with water and most organic solvents. It is the simplest aldol condensation product of acetone, and because of its keto-alcohol functionalities it has special utility in the coatings industry where it is used to dissolve cellulose acetate to give solutions with high tolerance for water (115).

Manufacture. Diacetone alcohol is manufactured by the low temperature liquid-phase self-condensation of acetone in the presence of a solid base catalyst. The reaction is exothermic (−14.6 kJ mol^{−1} = −3.49 kcal mol^{−1}) (116), and is equilibrium controlled (117). The unfavorable effect of increasing temperature on the equilibrium constant has been documented (116); the equilibrium concentration of diacetone alcohol is 23.1 wt % at 0°C, and declines to 9.1 wt % at 30°C. Although low temperatures favor diacetone alcohol formation, kinetic considerations require that commercial operation is conducted at 10–20°C. Either single or multistage catalyst beds can be used: single-stage conversion requires lower inlet temperatures, multistage conversion can cascade to progressively lower inlet temperatures (118). Residence times of 20–60 minutes are typically required.

Suitable catalysts include the hydroxides of sodium (119), potassium (76,120), calcium (121–125), and barium (126–130). Many of these catalysts are susceptible to alkali dissolution by both acetone and DAA and yield a crude product that contains acetone, DAA, and traces of catalyst. To stabilize DAA the solution is first neutralized with phosphoric acid (131) or dibasic acid (132). Recycled acetone can then be stripped overhead under vacuum conditions, and DAA further purified by vacuum topping and tailing. Commercial catalysts generally have a life of about one year and can be reactivated by washing with hot water and acetone (133). It is reported (134) that the addition of 0.2–2 wt % methanol, ethanol, or 2-propanol to a calcium hydroxide catalyst helps prevent catalyst aging. Research has reported the use of more mechanically stable anion-exchange resins as catalysts (135–137). The addition of trace methanol to the acetone feed is beneficial for the reaction over anion-exchange resins (138).

Early patents indicated that because water inhibits the aldol condensation mechanism, it was necessary to dry recycle acetone to less than 1% water (139–142). More recent reports demonstrate DAA production from waste acetone containing 10–50% water (143), and enhanced DAA production over anion-exchange resins using acetone feeds that contain 3–10% water (144,145).

Industrially, a selectivity to DAA of between 90–95% can be achieved (64). The principal by-products are mesityl oxide and acetone trimers. *sym*-Triacetone dialcohol [3682-91-5] can form by condensation of acetone with diacetone alcohol (116). Dehydration of *sym*-triacetone dialcohol can yield semiphorone [5857-71-6] (6-hydroxy-2,6-dimethyl-2-hepten-4-one), which may in turn ring close to form 2,2,6,6-tetramethyl-γ-pyrone [1197-66-6], or ultimately dehydrate to phorone [504-20-1] (2,6-dimethyl-2,5-heptadien-4-one) (146). Similarly, an unsymmetrical acetone trimer can also be formed which dehydrates to 2,4-dimethyl-2,4-heptadiene-6-one. These impurities complicate the high purity recovery of DAA, and are thought to be responsible for a yellow discoloration of DAA. The addition of dibasic acid (147) or nitrogen containing carboxylic or phosphonic acids (148) has been patented as refined product stabilizing agents.

Uses. Diacetone alcohol is a widely used solvent in the coatings industry where it finds application in hot lacquers which require high boiling components, and in brushing lacquers where its mild odor, blush resistance, and flow-out properties are desired. Diacetone alcohol is also a solvent for nitrocellulose, cellulose acetate, and epoxy resins.

In addition to DAA's use in the production of MIBK, DAA also finds use as a specialty reaction intermediate. Hydrogenation of DAA at 100°C and 30 MPa (83) yields hexylene glycol (\$1.43/kg, October 1994), widely used in castor oil-based hydraulic brake fluids and as a solvent. Reaction of *p*-phenetidine [156-43-4] with DAA synthesizes Monsanto's Santoquin (ethoxyquin) [91-53-2] (149), an antioxidant used in animal feeds and also as a rubber additive. Diacetone alcohol (acetone-free) was available at \$1.32/kg as of October 1994.

Diisobutyl Ketone. Diisobutyl ketone (DIBK) (2,6-dimethyl-4-heptanone) is a colorless stable liquid with a peppermint odor. Some physical properties are listed in Table 1.

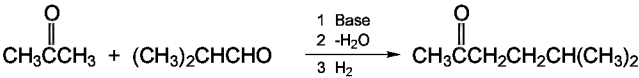
DIBK can be produced by the hydrogenation of phorone which, in turn, is produced by the acid-catalyzed aldol condensation of acetone. It is also a by-product in the manufacture of methyl isobutyl ketone. Diisobutyl ketone (\$1.37/kg, October 1994) is produced in the United States by Union Carbide (Institute, West Virginia) and Eastman (Kingsport, Tennessee) (47), and is mainly used as a coating solvent. Catalytic hydrogenation of diisobutyl ketone produces the alcohol 2,6-dimethyl-4-heptanol [108-82-7].

Methyl Isopropyl Ketone. Methyl isopropyl ketone [563-80-4] (3-methyl-2-butanone) is a colorless liquid with a characteristic odor of lower ketones. It can be produced by hydrating isoprene over an acidic catalyst at 200–300°C (150,151) or by acid-catalyzed condensation of methyl ethyl ketone and formaldehyde to 2-methyl-1-buten-3-one, followed by hydrogenation to the product (152). Other patented preparations are known (155,156). Methyl isopropyl ketone is used as an intermediate in the production of pharmaceuticals and fragrances (see PERFUMES), and as a solvent (157). It is domestically available from Eastman (Longview, Texas) (47).

Methyl Amyl Ketone. Methyl amyl ketone [110-43-0] (MAK) (2-heptanone) is a colorless liquid with a faint fruity (banana) odor. It is found in oil of cloves and cinnamon-bark oil, and is manufactured by the condensation of acetone and butyraldehyde (158). Other preparations are known (159–162).

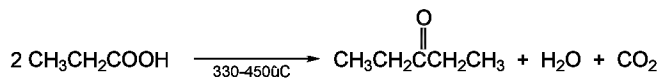
MAK is used as a high solids coating solvent (163) and in fragrances. It is available in the United States from Eastman (Kingsport, Tennessee), International Chemical Group (La Mesa, California) (47), and Union Carbide (South Charleston, West Virginia), and was priced at \$1.36/kg in October 1994.

Methyl Isoamyl Ketone. Methyl isoamyl ketone [110-12-3] (5-methyl-2-hexanone) is a colorless liquid with a mild odor. It is produced by the condensation of acetone and isobutyraldehyde (164) in three steps which proceed via the keto-alcohol dehydration to 5-methyl-3-hexen-2-one, and hydrogenation to 5-methyl-2-hexanone.



Isobutyraldehyde is commonly available as a by-product of propylene Oxo hydroformylation. Methyl isoamyl ketone is used as a solvent for cellulose esters, acrylics, and vinyl polymers. It is available in the United States from Eastman (Kingsport, Tennessee) (47) and Union Carbide (South Charleston, West Virginia) and was priced at \$1.42/kg in October 1994.

Diethyl Ketone. Diethyl ketone [96-22-0] (3-pentanone) is isomeric with methyl *n*-propyl ketone (2-pentanone), which has similar solvent and physical properties. Diethyl ketone is produced by the decarboxylation of propionic acid over MnO₂-alumina (165), ZrO₂ (166), or ZrO₂ or ThO₂ on TiO₂ (167,168). Diethyl ketone can also be produced by the hydrocarbonylation of ethylene (169–171). It is used as a solvent and a reaction intermediate.



Unsaturated Ketones

Mesityl Oxide. Mesityl oxide (MSO) (4-methyl-3-penten-2-one) is an oily colorless liquid with an unpleasant odor. It exhibits the versatility and unusual reactivity associated with conjugated α,β -unsaturated carbonyl compounds (172). On standing in air, mesityl oxide slowly forms bis(3,5,5-trimethyl-1,2-dioxolanyl)-3-peroxide (173).

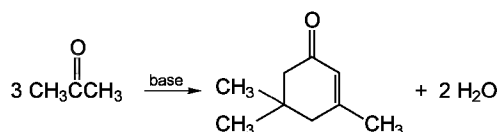
Commercial mesityl oxide can contain 5–20% of the β,γ -unconjugated isomer isomesityl oxide [141-79-7] (4-methyl-4-penten-2-one). At equilibrium, the mixture contains 91% of the α,β -mesityl oxide and 9% of the β,γ -isomer (174–176). Equilibrium is catalyzed by either acid or alkali. Techniques to isolate the isomers have been reported (174,176), and some physical properties of isomesityl oxide are reported in Table 1.

Manufacture. Mesityl oxide is produced by the liquid-phase dehydration of diacetone alcohol in the presence of acidic catalysts at 100–120°C and atmospheric pressure. As a precursor to MIBK, mesityl oxide is prepared in this manner in a distillation column in which acetone is removed overhead and water-saturated mesityl oxide is produced from a side-draw. Suitable catalysts are phosphoric acid (177,178) and sulfuric acid (179,180). The kinetics of the reaction over phosphoric acid have been reported (181).

Mesityl oxide can also be produced by the direct condensation of acetone at higher temperatures. This reaction can be operated in the vapor phase over zinc oxide (182), or zinc oxide–zirconium oxide (183), or in the liquid phase over cation-exchange resin (184) or zirconium phosphate (185). Other catalysts are known (186).

Health and Safety Factors. Mesityl oxide is more toxic than saturated ketones and is highly irritating to all tissues on vapor or liquid contact and for this reason sales of mesityl oxide ceased in the United States in 1986. It is absorbed through intact skin, and prolonged exposure can damage liver, kidneys, and lungs. Repeated exposure to vapors can cause anemia and leukopenia (187); however, the odor is so intolerable that long-term exposure is unlikely. Mesityl oxide is still produced, but is consumed captively as an intermediate in the production of MIBK, methyl isobutyl carbinol, and isophorone.

Isophorone. Isophorone (3,5,5-trimethyl-2-cyclohexen-1-one) is a cyclic α,β -unsaturated ketone derived from the trimerization of acetone. It has a light yellow color and a disagreeable camphoraceous odor. It has the tendency to discolor and form residues on prolonged storage. Isophorone is completely miscible with organic solvents, and other physical properties are listed in Table 1.



Isophorone usually contains 2–5% of the isomer β -isophorone [471-01-2] (3,5,5-trimethyl-3-cyclohexen-1-one). The term α -isophorone is sometimes used in referring to the α,β -unsaturated ketone, whereas β -isophorone connotes the unconjugated derivative. β -Isophorone (bp 186°C) is lower boiling than isophorone and can be converted to isophorone by distilling at reduced pressure in the presence of *p*-toluenesulfonic acid (188). Isophorone can be converted to β -isophorone by treatment with adipic acid (189) or iron(III) acetylacetoate (190). β -Isophorone can also be prepared from 4-bromoisophorone by reduction with chromous acetate (191). β -Isophorone can be used as an intermediate in the synthesis of carotenoids (192).

Manufacture. Isophorone is produced by aldol condensation of acetone under alkaline conditions. Severe reaction conditions are required to effect the condensation and partial dehydration of three molecules of acetone, and consequently raw material inefficiency to by-products is limited by employing low conversions. Both liquid- and vapor-phase continuous technologies are practiced (186,193,194).

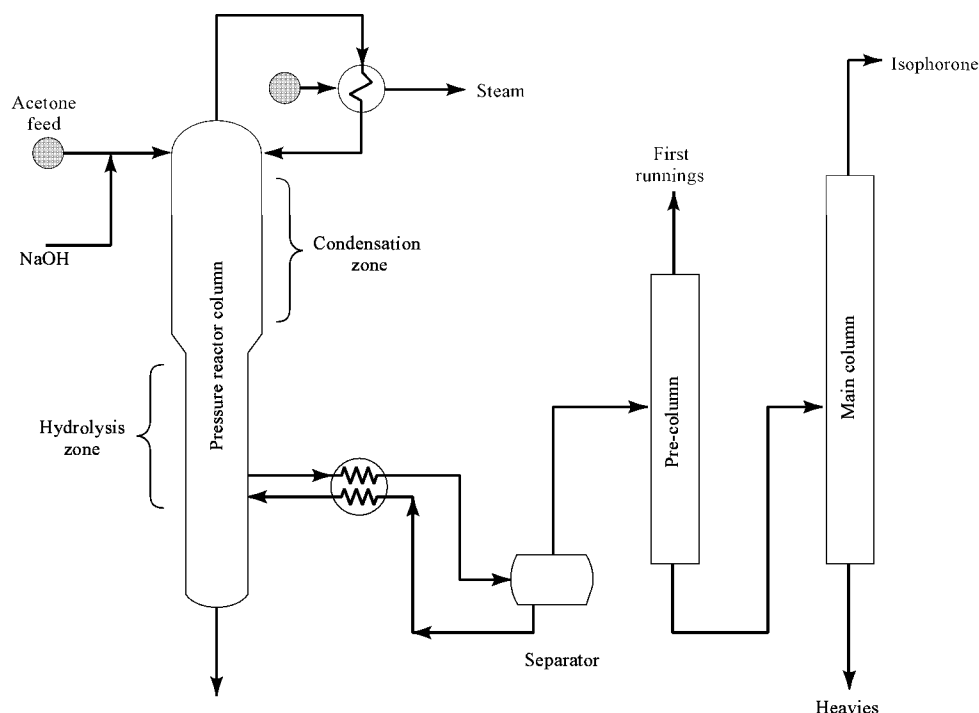


Fig. 4. Liquid-phase isophorone process (83).

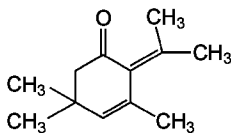
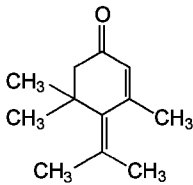
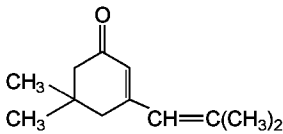
Courtesy of Chemische Industrie.

A liquid-phase isophorone process is depicted in Figure 4 (83). A mixture of acetone, water, and potassium hydroxide (0.1%) are fed to a pressure column which operates at head conditions of 205°C and 3.5 MPa (~500 psi). Acetone condensation reactions occur on the upper trays, high boiling products move down the column, and unreacted acetone is distilled overhead in a water–acetone azeotrope which is recycled to the column as reflux. In the lower section of the column, water and alkali promote hydrolysis of reaction by-products to produce both isophorone and recyclable acetone. Acetone conversion is typically in the range 6–10% and about 70% yield of isophorone is obtained. Condensation–hydrolysis technology (195–198), and other liquid-phase production processes have been reported (199–205).

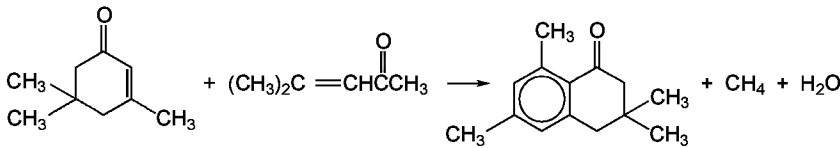
In the vapor phase, acetone vapor is passed over a catalyst bed of magnesium aluminate (206), zinc oxide–bismuth oxide (207), calcium oxide (208), lithium or zinc-doped mixed magnesia–alumina (209), calcium on alumina (210), or basic mixed-metal oxide catalysts (211–214). Temperatures ranging

from 250–400°C and space velocities of 0.5–1.0 liquid hourly space velocity (LHSV) are employed.

The liquid-phase processes are more energy efficient than the vapor-phase processes, however, they incur costly high pressure equipment investment and also produce waste streams containing used catalyst (213). Both methods produce substantial quantities of by-products which cause refining difficulties. The by-products consist primarily of mesitylene [108-67-8], phorone [504-20-1], and the following xylitone isomers (215):



Substantial amounts of 3,3,6,8-tetramethyl-1-tetralone [5409-55-2] are also formed, most notably in the vapor-phase process (216). This tetralone has been synthesized from isophorone and mesityl oxide and it can thus be assumed to be a product of these two materials in the isophorone process (217,218).



Some reversion of the over-condensate residues to acetone and isophorone is possible by hydrolysis with 2% sodium hydroxide solution at 175°C and 0.9 MPa (219).

The crude isophorone refining operation can require as many as three or four distillation columns, and is often conducted under vacuum to limit further by-product formation at high temperatures. Discoloration treatment during refining is often required to meet coatings industry requirements and to prolong storage life; various treatments using *p*-toluenesulfonic acid (198,220), oxalic acid (221), acidified fuller's earth (222), diazines (223), diisopropylamine (224), polyhydroxybenzene (225), aqueous caustic (226), and ion exchange (227,228) have been patented. Refined isophorone (99%) typically has a platinum–cobalt value of about 30–50 (APHA).

Economic Aspects. Isophorone was available at \$1.87/kg in October 1994. The sole domestic producer of isophorone is Union Carbide; however, Huls is by far the largest isophorone producer in the world. Other significant producers are listed in Table 9. Despite the erosion of some of the historical solvent uses of isophorone, the expanding derivatives market for this product appear to sustain its production in the short term.

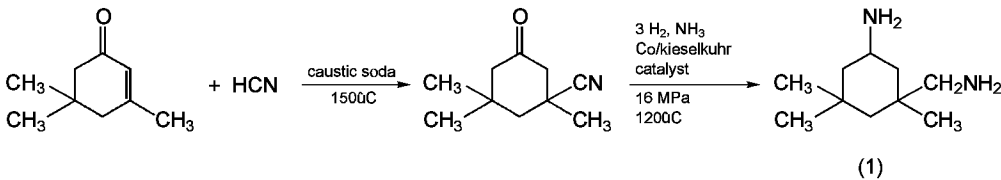
Table 9. World Isophorone Producers^a

Producer	Plant location	Process	1992 Capacity, 10 ³ t/yr
Huls	Herne, Germany	alkali liquid phase	44
Atochem	La Chambre, France	alkali liquid phase	10
SpA - SISAS	Piotello, Italy	na	9
British Petroleum	Hull, England	alkali liquid phase	6
Union Carbide	Institute, West Virginia	vapor phase	na
Total			~78

^a Ref. 47.
Courtesy of SRI International.

Uses. Isophorone has traditionally been used predominantly as a high boiling, low evaporative solvent. It exhibits powerful solvation power for a large number of natural and synthetic polymers, resins, fats, and oils. Of note are its use as a solvent for formulating highly concentrated vinyl chloride–acetate-based coating systems for metal cans, nitrocellulose finishes, and as a leveling aid to prevent blistering and promote flow for uniform wetting of metal paints based on polyacrylates, alkyds, and epoxy and phenol–formaldehyde resins (229). In addition, it is used as a solvent for insecticide and herbicide concentrates where it is employed to produce high emulsibility and good emulsion stability in aqueous dilution (83) (see HERBICIDES; INSECT CONTROL TECHNOLOGY). Exceptions which isophorone does not dissolve are polyethylene, polypropylene, polyamides, and polyurethanes (229). Isophorone has also been suggested as an effective woodpecker repellant when coated onto telegraph poles.

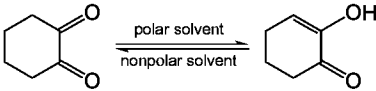
A trend in the utility of isophorone is as an important industrial building block. Foremost among these developments has been the use of isophorone as a raw material, and isophorone diisocyanate [2855-13-2] (IPDI), for the production of the light-stable polyurethane. The U.S. market for IPDI-based products was \$31 million in 1989, and is estimated to grow to \$53 million in 1994 (230).



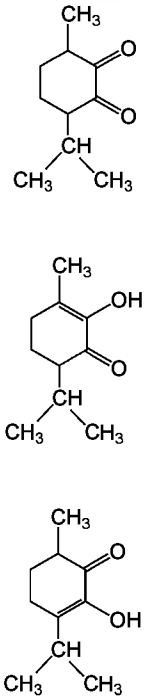
3,4-heptanedione	[13706-89-3]	propionylbutyrl	147 (98)		885.0 ⁰		
5-methyl-2,3- hexanedione	[13706-86-0]	acetylisovaleryl	138	1.4119 ²⁰	908.0 ²²		432
2,3-octanedione	[585-25-1]	acetylcaproyl	172–173 (98)				
4,5-octanedione	[5455-24-3]	bibutyryl	168		934.0 ⁰	yellow	435
2,5-dimethyl-3,4-heptanedione	[4388-87-8]	biisobutyryl	144–145	1.4206 ²⁰	923.2 ²⁰		436
5-methyl-3,4- heptanedione	[13678-56-3]		63–67 (5.3)				
6-methyl-3,4- heptanedione	[3131-90-6]		53–54 (2.0)	1.4151 ²⁰	901.9 ²⁰		
			Cyclic				
1,2-cyclopentanedione	[3008-40-0]		105 (2.7)	55–56			
1,2-cyclohexanedione	[765-87-7]		193–195	35–38	1.4995 ²⁰		
			Aromatic				
benzil	[134-81-6]	bibenzoyl	346–348	94–95		yellow	
			188 (1.6)				
1-phenyl-1,2-propanedione	[579-07-7]	acetylbenzoyl	125 (3.1)		1101 ²⁰		
1,2-naphthalendione	[524-42-5]	1,2-naphtho-quinone		145–147 (dec)		golden yellow	

^a Refs. 282–r285.

Cyclic 1,2-diketones demonstrate enolic tautomerism, with solvent polarity affecting tautomeric equilibrium:



Diosphenol [490-03-9], the main constituent of buchu leaves, is an example of a naturally occurring compound with tautomeric properties (286):



1,2-Diketones can be prepared by oxidation of the corresponding monoketone (287) or α-hydroxyketone (288). 1,2-Diketones are used extensively as intermediates in the preparation of pharmaceuticals, flavors, and fragrances. Toxicity data for selected diketones are shown in Table 11.

Table 11. Toxicological Properties of Diketones^a

Ketone	Ingestion, LD ₅₀ ^a , rats, mg kg	Inhalation, LC _{LO} ^b , rats, ppm 4 h	Injection into the peritoneal cavity	
			LD ₅₀ ^a , mg kg	LD _{LO} ^c , mg kg
1,2-diketones				
benzil	2710			
biacetyl	1580		400 ^d	
1,3-diketones				
2,4-pentanedione	1000	1000	750 ^e	400 ^d
1,3-cyclohexanedione				64 ^e
1,4-diketones				
2,5-hexanedione	2700	2000		
1,4-cyclohexanedione				100 ^d

^a Ref. 289.

^b The lowest concentration in air which caused death in four hours.

^c The lowest dose resulting in death.

^d Rat.

^e Mouse.

Biacetyl. Biacetyl [431-03-8] (2,3-butanedione) is a greenish yellow liquid with a quinone odor. Biacetyl occurs naturally in bay oil and is readily soluble in organic solvents. It is a constituent of many food aromas, eg, butter, and is commonly used to flavor margarine. Flavor-grade biacetyl was available at \$20.40 kg in July 1993, and is used as an odorant for coffee, vinegar, tobacco, and in perfumes.

Biacetyl is produced by the dehydrogenation of 2,3-butanediol with a copper catalyst (290,291). Prior to the availability of 2,3-butanediol, biacetyl was prepared by the nitrosation of methyl ethyl ketone and the hydrolysis of the resultant oxime. Other commercial routes include passing vinylacetylene into a solution of mercuric sulfate in sulfuric acid and decomposing the insoluble product with dilute hydrochloric acid (292), by the reaction of acetal with formaldehyde (293), by the acid-catalyzed condensation of 1-hydroxyacetone with formaldehyde (294), and by fermentation of lactic acid bacterium (295–297). Acetoin [513-86-0] (3-hydroxy-2-butanone) is also coproduced in lactic acid fermentation.

Benzil. Benzil [134-81-6] (diphenylethanedione) is a yellow solid that crystallizes from alcohol in hexagonal prisms. Benzil can be prepared by the oxidation of benzoin [579-44-2] (2-hydroxy-2-phenylacetophenone) (298,299), which is itself prepared by the self-condensation of benzaldehyde (300). Benzil is commercially produced in Japan and is used as a uv resin curing sensitizer (301). It has also been suggested as a chigger repellent (302).

1,3-Diketones. β-Diketones contain two carbonyl groups separated by one carbon atom. Aliphatic 1,3-diketones are colorless liquids whose boiling point increases with increasing molecular weight, whereas cyclic 1,3-diketones are colorless solids. 1,3-Diketones are miscible with organic solvents, and lower group members have some water solubility. Many of the higher members of the series possess ester-like odors and are used in fragrances (303): 3-methyl-2,4-heptanedione [13152-54-0] has a fruity aroma, tetramethylcyclobutanedione [29714-52-3] a minty smell, and benzoylacetone [93-91-4], a balsam-like odor. Some physical properties of 1,3-diketones are shown in Table 12.

Table 12. Physical Properties of 1,3-Diketones^a

CAS name	CAS Registry Number	Synonyms	Bp, °C ^b	Melting point, °C	Density, g L
<i>Aliphatic</i>					
2,4-hexanedione	[3002-24-2]	propionylacetone	158		959
3,5-heptanedione	[7424-54-6]	dipropionylmethane	47 (0.8)		944.5
2,4-heptanedione	[7307-02-0]	butyrylacetone	174–175		941.1 ^c
3,5-octanedione	[6320-18-9]	butyrylpropionyl-methane	189–190		
5-methyl-2,4-hexanedione	[7307-03-1]	isobutyrylacetone	168		
2,6-dimethyl-3,5-heptanedione	[18362-64-6]	diisopropionyl-methane	66 (1.1)		
2,4-octanedione ^d	[14090-87-0]	valerylacetone	79–83 (2.7)		923.3
5,5-dimethyl-2,4-hexanedione	[29284-62-6]	pivaloylacetone	164–167 (99)		
6-methyl-2,4-heptanedione	[3002-23-1]	isovalerylacetone	73 (2.4)		
<i>Cyclic</i>					
1,3-cyclopenta-nedione	[3859-41-4]			151–153	
1,3-cyclohexa-nedione	[504-02-9]	dihydroresorcinol		103–105	
5,5-dimethyl-1,3-cyclohexa-nedione	[126-81-8]	dimedone, methone		149–151	
<i>Aromatic</i>					
1-phenyl-1,3-butanedione	[93-91-4]	benzoylacetone	98–100 (0.3)	58–60	1090
1-phenyl-1,3-pentanedione	[5331-64-6]	benzoylpropionyl-methane	92–94 (0.13)		
1,3-diphenyl-1,3-propanedione	[120-46-7]	dibenzoylmethane	219–221	77.5–79	
1-phenyl-2,4-pentanedione ^e	[3318-61-4]		144		

^a Refs. 282, 283, 304, 305.
^b At 101.3 kPa (760 mm Hg) unless otherwise indicated in parentheses, kPa.
^c At 15°C.
^d n_D^{20} 1.4559.
^e n_D^{20} 1.5837.

Because the structure of 1,3-diketones comprise a methylene group between two activating carbonyls, equilibrium is shifted toward the enol form. The equilibrium distribution varies with structure and solvent (303,306) (Table 13). The enol forms are cyclic and acidic and form covalent, colored, solid chelates with metals:

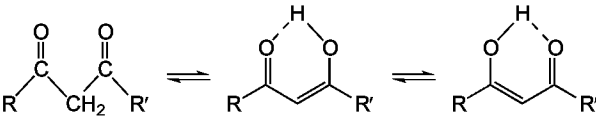


Table 13. Enol and Chelating Properties of 1,3-Diketones

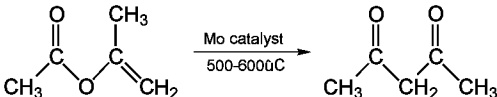
1,3-Diketone	Enol, %	Refractive index	Copper chelate	
			Color	Mp, °C
		Aliphatic		
2,4-pentanedione	76.4			
2,4-hexanedione	80.2	1.4516 ^a	green	198
3,5-heptanedione				209–214
2,4-heptanedione	83.6		blue	161
3,5-octanedione			blue	158
5-methyl-2,4-hexanedione				171
		Cyclic		
1,3-cyclohexane-dione	100			

		Aromatic		
1-phenyl-1,3-butanedione	99		sea green	199
1-phenyl-1,3-pentanedione		1.5731 ^b	sea green	135
1,3-diphenyl-1,3-propanedione			green	296–300
^a n_D^{20} .				
^b n_D^{25} .				

1,3-Diketones are used for extraction and identification of metals, and as raw materials for synthesis of heterocyclic compounds. The ability of 1,3-diketones to tie up essential metal ions in animal systems causes 1,3-diketones to be more toxic than other diketones.

2,4-Pentanedione. 2,4-Pentanedione [123-54-6] (acetylacetone) is the lowest member of the aliphatic 1,3-diketones and is a colorless liquid with a mild ketone-like odor. It is completely miscible with organic solvents; other physical properties are shown in Table 1.

The industrial precursor to 2,4-pentanedione is isopropenyl acetate, produced from acetone and ketene (307,308). The diketone is formed by the high temperature isomerization of isopropenyl acetate over a metal catalyst (309–311).



2,4-Pentanedione can also be produced by the condensation of acetone with ethyl acetate (312–317), or by the condensation of ethyl acetoacetate and ketene (318–321). Other methods are known (322,323).

2,4-Pentanedione is widely used in extraction processes for the separation and purification of metals because of its ability to form covalent metal chelates. It is also used as an intermediate in the production of heterocyclic substances and dyes, as a fuel additive (324), and in metal plating and resin modification.

The toxicity of 2,4-pentanedione is shown in Tables 3 and 11 to be similar to mesityl oxide, and greater than most other 1,2- or 1,4-diketones or monoketones. Inhalation of low levels of 2,4-pentanedione can cause nausea, eye contact can induce stinging, and recurrent exposure to high concentrations (300–400 ppm) can adversely affect the central nervous system and immune system (325).

1,4-Diketones. γ -Diketones contain two carbonyl groups separated by two carbon atoms. With the exception of 2,5-hexanedione which is a high boiling liquid, 1,4-diketones are low melting white solids with only faint odors. Lower members are soluble in organic solvents and water. Properties of representative 1,4-diketones are shown in Table 14.

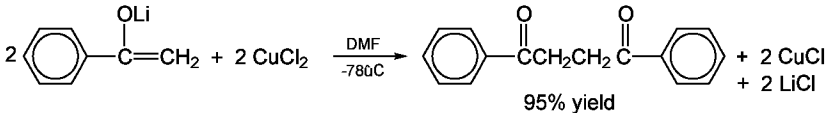
Table 14. Physical Properties of 1,4-Diketones^a

CAS name	CAS Registry Number	Common name	Bp, °C (kPa) ^b	Melting point, °C	Refractive index, n_D^{20}
3,4-dimethyl-2,5-hexanedione	[25234-79-1]		92 (4.0)		1.4330
3,3,4,4-tetra-methyl-2,5-hexanedione	[23328-38-3]		40 (0.7)		1.4522
2,5-heptanedione	[1703-51-1]	acetylpropionyl-ethane	90 (2.8)		
3,6-octanedione	[2955-65-9]	dipropionylethane	98 (1.9)	34–35	
6-methyl-2,5-heptanedione	[13901-85-4]		91 (1.6)		
2,5-decanedione	[41368-32-5]		132 (2.3)		
2,5-dodecane-dione	[32781-66-1]		148 (1.6)	40.5	
1,4-cyclohexane-dione	[637-88-7]			77–78.5	
1,4-diphenyl-1,4-butanedione	[495-71-6]	1,2-dibenzoyl-ethane		145–147	

^a Refs. 282 and 326.

^b To convert kPa to mm Hg, multiply by 7.5.

1,4-Diketones are intermediates for synthesis of perfumes and natural products, and several preparative methods have been developed (327); in the simplest preparative methods, ketone enolates are oxidatively dimerized (328):



1,4-Diketones are readily transformed to cyclic derivatives, such as cyclopentanones and furans. In this manner, the fragrance dihydrojasnone (3-methyl-2-pentyl-2-cyclopenten-1-one) is prepared by the base-catalyzed aldol condensation of 2,5-undecanedione. 2,5-Undecanedione is itself prepared from heptanal and methyl vinyl ketone in the presence of thiazolium salts (329). *cis*-Jasmone can be similarly prepared (330,331).

2,5-Hexanedione [110-13-4] (acetylacetone) is one of the most widely used 1,4-diketones. It is a colorless high boiling liquid prepared by the hydrolysis of 2,5-dimethylfuran (332,333), by oxidation of 2,5-hexanediol (334) or 5-hexen-1-one (335), and from allylacetone (336). Its main use is in solvent systems and as a raw material for chemical synthesis. It is reportedly not highly toxic (336).

Cyclic Ketones

Cycloaliphatic ketones are colorless liquids with boiling points that increase regularly with increasing molecular weight. Virtually all members of the series exhibit a characteristic odor depending on the ring size (337). Physical properties are given in Table 15.

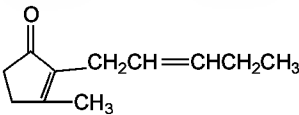
Ring size	Odor
5	bitter almonds
6	peppermint
7–9	transition to camphor-like
10–12	camphor-like
13	cedar wood-like
14	musk-like

Table 15. Properties of Cyclic Ketones,^a $\boxed{\text{(CH}_2\text{)}_x\text{C=O}}$

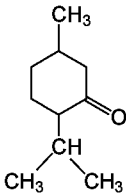
CAS name	CAS Registry Number	Formula x n	Bp, °C ^b	Melting point, °C	Refractive index, n_D^{20}	Density at 20°C, ^c g L	C=O stretching frequency, max, cm ⁻¹	Enol, %
cyclopropanone	[5009-27-8]	2	d					
cyclobutanone	[1191-95-3]	3	100–102		1.4195	938	1788	0.55
cyclopentanone	[120-92-3]	4	130		1.4359	951	1746	0.09
cyclohexanone	[108-94-1]	5	156.7	47	1.4507	948		
cycloheptanone	[502-42-1]	6	179–181		1.4611	951	1703	0.56
cyclooctanone	[502-49-8]	7	74 (1.6)	42		958	1703	9.3
cyclononanone	[3350-30-9]	8	93–95 (1.6)	34	1.4770	959	1702	4.0
cyclodecanone	[1502-06-3]	9	107 (1.7)	29	1.4820	958		6.1
cycloundecanone	[878-13-7]	10	108 (1.6)	10	1.4804			
cyclododecanone	[830-13-7]	11	125 (1.6)	61		906		
cyclotridecanone	[832-10-0]	12	138 (1.6)	32	1.4790	927		
cyclotetradecanone	[3603-99-4]	13	155 (1.6)	53				
cyclopentadecanone	[502-72-7]	14	120 (0.04)	63		897		
cyclohexadecanone	[2550-52-9]	15	138 (0.04)	60				
cycloheptadecanone	[3661-77-6]	16	141 (0.13)	63				
cyclooctadecanone	[6907-37-5]	17	158 (160)	72				
cyclononadecanone	[6907-38-6]	18	160 (0.04)	72				
cyclocosanone	[6907-39-7]	19	171 (0.04)	59				

^a Refs. 283, 338–340.
^b At 101.3 kPa (760 mm Hg) unless otherwise indicated in parens, kPa.
^c For solids, the density is given for the liquid at the melting point temperature.
^d Rapidly polymerizes at room temperature. A stable hydrate, mp 71–72°C, is formed in water.

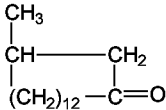
Many cyclic ketones occur in natural oils. Jasmone [488-10-8] (3-methyl-2-(2-pentyl)-2-cyclopenten-1-one) (4) is an odoriferous component of the oil obtained from jasmine flowers. *l*-Menthone [14073-97-3] (5) is the most frequently occurring of four optically active isomers, and is a colorless liquid with a minty odor obtained from *Mentha* species of plants. Muscone [541-91-3] (6) and civetone [542-46-1] (7) are expensive animal products.



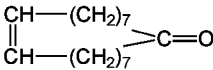
(4)



(5)

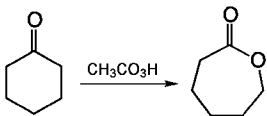


(6)



(7)

The chemical properties of cyclic ketones also vary with ring size. Lower members (<C₈) are more reactive, for example, toward addition reactions, than corresponding acyclic ketones. The C₈–C₁₂ ketones are unreactive, reflecting the strain and high enol content of medium-sized ring systems. Lactones are prepared from cyclic ketones by the Bayer-Villiger oxidation reaction with peracids. ε-Caprolactone is manufactured from cyclohexane by this process:



Some toxicological data for cyclic ketones are shown in Table 16. Interestingly, toxicity is shown to increase with ring size. This is in reverse order of relative reactivity.

Table 16. Toxicological Properties of Cyclic Ketones Compared with Aliphatic Ketones^a

Ketone	Administered into the peritoneal cavity	Administered under the skin LD ₅₀ ^b
	LD ₅₀ , mouse, mg kg	mouse, mg kg
cyclopentanone	1950	2600
cyclohexanone	1350	1300
cycloheptanone	750	930
cyclooctanone	740	
acetone	1297	
mesityl oxide	268	
methyl ethyl ketone	626	

^a Ref. 290.

^b The lowest lethal dose.

Cyclohexanone is by far the most important industrial cyclic ketone (see CYCLOHEXANOL AND CYCLOHEXANONE).

Aromatic Ketones

Aromatic ketones of industrial significance include acetophenone, propiophenone, and benzophenone.

Acetophenone. Acetophenone [98-86-2] (methyl phenyl ketone) is a colorless liquid that forms laminar crystals at low temperature (mp 20°C). It has a characteristic sweet orange blossom odor, and is soluble in alcohols and ethers. It is found in nature in oil of casatorem, obtained from beavers; oil of labdanum, recovered from plants; and in buds of balsam poplar. It can be prepared by the Friedel-Crafts reaction (qv) of acetyl chloride with benzene in the presence of aluminum chloride; however, this route is of little commercial significance.

Sales demand for acetophenone is largely satisfied through distillative by-product recovery from residues produced in the Hock process for phenol (qv) manufacture. Acetophenone is produced in the Hock process by decomposition of cumene hydroperoxide. A more selective synthesis of acetophenone, by cleavage of cumene hydroperoxide over a cupric catalyst, has been patented (341). Acetophenone can also be produced by oxidizing the methylphenylcarbinol intermediate which is formed in styrene (qv) production processes using ethylbenzene oxidation, such as the ARCO and Halcon process and older technologies (342,343).

Acetophenone can react with formaldehyde to yield light-resistant resins which are used as additives in nitrocellulose paints. It is also used as a photoinitiator, and in the pharmaceuticals, perfumery, and pesticide industries (344). It can be hydrogenated to 1-phenylethanol which is used for the production of aromatic ester fragrances (345). Technical-grade acetophenone is available at \$2.29 kg; perfume-grade acetophenone was \$6.50 kg in October 1994.

Propiophenone. Propiophenone [93-55-0] (ethyl phenyl ketone) is a colorless liquid with a flowery odor. It can be prepared by the Friedel-Crafts reaction of benzene and propionyl chloride in the presence of aluminum chloride (346), or by the catalytic reaction of benzoic acid and propionic acid in the presence of water (347). Propiophenone is commercially available (348), and is sold in Japan at 2700 Y kg (349). It is used in the production of ephedrine, as a fragrance enhancer, and as a polymerization sensitizer.

Benzophenone. Benzophenone [119-61-9] (diphenyl ketone) exists in a stable form as colorless orthorhombic bisphenoidal prisms when crystallized from alcohol or ether. Other labile forms of lower melting point exist. Benzophenone has been identified as a flavor component of wine grapes and has a geranium-like odor. It is soluble in most organic solvents, and is insoluble in water.

Benzophenone is produced by the oxidation of diphenylmethane (350). This free from chlorine (FCC) route is favored for perfume uses. The Friedel-Crafts reaction of benzene and benzoyl chloride in the presence of aluminum chloride is also possible; this reaction may proceed in the absence of catalyst at a temperature of 370°C and pressure of 1.4 MPa (351).

Benzophenone is used as a photoinitiator in ultraviolet-curable printing inks, coatings, and adhesive formulations; as a uv-light absorbing agent in personal care products; and as a perfume and flavor enhancer. It is also used as an intermediate for agricultural and pharmaceutical chemicals, eg, in the production of benzophenone hydrazone, a blocking agent used in the manufacture of penicillin (352). It was commercially available in the United States (353,354) at \$5.30 kg (tech-flakes), \$5.51 kg (crystals), \$8.05 kg (flakes) in July 1993.

BIBLIOGRAPHY

"Ketones" in *ECT* 1st ed., Vol. 8, pp. 113–151, by H. J. Hagemeyer, Jr., Tennessee Eastman Co., P. R. Rector, Carbide and Carbon Chemical Co., C. O. Blackburn, Shell Development Co., and G. A. Reynolds, M. Ream, and J. A. Von Allen, Eastman Kodak Co.; in *ECT* 2nd ed., Vol. 12, pp. 101–169, by A. P. Lurie, Eastman Kodak Co.; in *ECT* 3rd ed., Vol. 13, pp. 894–941, by A. J. Papa and P. D. Sherman, Jr., Union Carbide Corp.

1. S. R. Sandler and W. Karo, *Organic Functional Group Preparations*, Vol. 1, 2nd ed, Academic Press, Inc., New York, 1983.
2. Physical property data, Union Carbide Corp. Danbury, Conn.
3. R. W. Gallant, *Hydrocarbon Proc.* **47**, 8, 127 (1968).
4. F. A. Patty, in G. D. Clayton and F. E. Clough, eds., *Industrial Hygiene and Toxicology*, John Wiley & Sons, Inc., New York, 1991.
5. R. J. Lewis, Sr., *Sax's Dangerous Properties of Industrial Materials*, 8th ed., Van Nostrand Reinhold, New York, 1992.
6. S. Zakhari and co-workers, *Isopropanol and Ketones in the Environment*, CRC Press, Cleveland, Ohio, 1977.
7. *Code of Federal Regulations*, Title 40, Section 63.185, List of Volatile Hazardous Air Pollutants, Washington, D.C., 1990.
8. *Chem. Week*, 9 (Oct. 17, 1990).
9. P. W. Fisher, R. M. Currie, and R. J. Churchill, *J. Air Poll. Control Assoc.* **38**, 1376 (1988).
10. *Chem. Mark. Rep.* "Coatings '92 Special Report, SR22," (Oct. 19, 1992).
11. J. J. Perona, and G. Thodos, *AIChEJ* **3**, 230 (1957).
12. L. H. Thaller and G. Thodos, *AIChEJ* **6**, 369 (1960).
13. F. E. Ford and D. D. Perlmutter, *Chem. Eng. Sci.* **19**, 371 (1964).
14. H. J. Kolb and R. L. Burrell, *J. Am. Chem. Soc.* **67**, 1084 (1945).
15. E. Matsunaga and R. G. Muller, *Secondary Butyl Alcohol via Direct Hydration*, Process Economics Program, Review No. 84-2-2, SRI International, Menlo Park, Calif., Aug. 1985.

16. **Brit. Pat. 1,368,195** (Mar. 5, 1975), (to Deutsche Texaco).

17. **Brit. Pat. 1,381,455** (Jan. 22, 1975), G. Brandes and co-workers (to Deutsche Texaco).

18. **U.S. Pat. 4,340,769** (July 20, 1982), G. Brandes and co-workers (to Deutsche Texaco).

19. **U.S. Pat. 4,476,333** (Oct. 9, 1984), W. Neier and co-workers (to Deutsche Texaco).

20. **Brit. Pat. 665,376** (Jan. 23, 1952), (to Standard Oil Development).

21. **U.S. Pat. 2,835,706** (May 20, 1958), C. E. Cordes (to ESSO Research and Engineering).

22. **U.S. Pat. 2,885,442** (May 5, 1959), McCulloch and co-workers (to ESSO Research and Engineering).

23. **Brit. Pat. 779,350** (July 17, 1957), (to Rheinpreussen).

24. **Rus. Pat. 43-3,163** (Feb. 5, 1968), T. Miyata (to Toyo Rayon).

25. DE Pat. 1,130,800 (June 7, 1962), F. Engelhardt (to Rheinpreussen).

26. DE Pat. 1,026,739 (Mar. 27, 1962), W. Opitz and co-workers (to Knapsack-Griesheim).

27. **Brit. Pat. 1,269,167** (Apr. 6, 1972), B. Stouthamer and A. Kwantes (to Shell).

28. Technical brochure, *The Maruzen MEK Process*, Maruzen Oil Co. Ltd., Aug. 1969.

29. S. Takaoka, *Acetone, Methyl Ethyl Ketone, and Methyl Isobutyl Ketone*, Process Economics Program, Report No. 77, SRI, Menlo Park, Calif., May 1972.

30. W. Neier, in W. Gerhartz, ed., *Ullmans Encyclopedia of Industrial Chemistry*, 5th ed., VCH Publishers, New York, 1985.

31. **U.S. Pat. 3,196,182** (July 20, 1965), N. R. Cox (to Union Carbide Corp.).

32. **U.S. Pat. 2,704,294** (Mar. 15, 1955), C. S. Morgan and co-workers (to Celanese Corp.).

33. J. B. Saunby and B. W. Kiff, *Hydrocarbon Proc.*, 247 (Nov. 1976).

34. J. Smidt and H. Krekeler, *Hydrocarbon Proc. Petrol. Refr.* **42**, 7, 149 (1963).

35. *Chem. Eng.* (Sept. 30, 1963).

36. *Hydrocarbon Proc.*, 204 (Nov. 1969).

37. **U.S. Pat. 3,236,897** (Feb. 22, 1966), L. Hoernig and co-workers (to Hoechst).

38. **U.S. Pat. 3,215,743** (Nov. 2, 1965), W. Riemenschneider (to Hoechst).

39. Neth. Pat. 68 17,336 (June 8, 1970), (to Maruzen).

40. *Chem. Mark. Rep.* **49** (July 28, 1986).

41. *Chem. Eng. News* **64**, 37, 20 (Aug. 4, 1986).

42. *Eur. Chem. News* **47**, 1237, 13 (Aug. 4, 1986).

43. *Chem. Eng.* **93**, 16, 24 (Aug. 18, 1986).

44. **U.S. Pat. 4,720,474** (Jan. 19, 1988), Vasilevskis and co-workers.

45. **U.S. Pat. 4,723,041** (Feb. 2, 1988), Vasilevskis and co-workers.

46. R. H. Schwaar and S. Morimoto, *Methyl Ethyl Ketone by Direct Oxidation of n-Butenes*, Process Economics Program, Review No. 87-2-3, SRI International, Menlo Park, Calif., Oct. 1988.

47. *Directory of Chemical Producers, Volumes for United States, East Asia, Western Europe*, SRI International, Menlo Park, Calif., 1992.

48. *Chem. Mark. Rep.* (Jan. 1, 1980).

49. *Chem. Mark. Rep.* (Aug. 27, 1990).

50. "ARCO to Exit MEK Business Citing Feedstock, Plant Age," *Chem. Mark. Rep.* (Jan. 7, 1991).

51. *Chem. Week.* **19** (Nov. 25, 1987).

52. *Chem. Mark. Rep.*, 9, 18, 19 (Oct. 17, 1988).

53. "Pampa Rebuilding on Target," *Chem. Mark. Rep.* (Nov. 21, 1988).

54. *J. Am. Paint Coatings* **72**, 54, 22 (1988).

55. *Chem. Week.*, 6 (Feb. 3, 1993).

56. L. Fujise, *Chemical Economics Handbook*, CEH Data Summary, SRI International, Menlo Park, Calif., July 1990.

57. *Chem. Week.*, 44 (Jan. 4 11, 1989).

58. P. W. Sherwood, *Erdol Kohle* **8**, 12, 884 (1955).

59. U.S. Pat. 3,374,272 (Mar. 19, 1969), A. Heykoop and F. A. Van Dijk (to Shell Oil).

60. *MIBK Direct Conversion of Acetone over Heterogeneous Catalyst-Sumitomo*, Process Evaluation Research Planning (PERP), Topical Reports, Vol. III, Chem Systems Inc., Tarrytown, NY, 1988.

61. C. S. Read and T. Gibson, *Chemical Economic Handbook*, CEH-Data Summary, SRI International, Menlo Park, Calif., 1990.

62. **Czech. Pat. 241,425** (Feb. 1, 1988), Z. Hejda and co-workers.

63. **Czech. Pat. 234,604** (Jan. 1, 1987), J. Pasek and co-workers.

64. *Hydrocarbon Proc.*, 205 (Nov. 1969).

65. **U.S. Pat. 3,362,822** (Jan. 2, 1968), K. Schmitt and co-workers (to Scholven-Chemie AG).

66. **Neth. Appl. 6,503,470** (Sept. 19, 1966), (to Shell Internationale Research Maatschappij NV).

67. **DE Pat. 1,234,665** (July 1967), A. Heykoop and F. A. Van Dijk (to Shell Internationale Research Maatschappij NV).

68. **Brit. Pat. 574,446** (Jan. 1946), C. Weizmann.

69. K. Hoshiai, *Kogyo Kakaku Zasshi* **60**, 1150 (1957).

70. K. P. Grinevich and V. A. Zaitsev, *Khim. Prom.*, 276 (1958).

71. **Neth. Appl. 6,403,176** (Oct. 28, 1964), (to Schloven AG).

72. **Fr. Pat. 1,478,704** (Apr. 28, 1967), B. W. Turnquest and O. H. Thomas (to Sinclair Research Inc.).

73. **Brit. Pat. 1,116,037** (June 6, 1968), L. J. Sirois and co-workers (to Esso Research and Engineering).

74. **Neth. Appl. 6,609,628** (Jan. 10, 1967), (to Esso Research and Engineering).

75. F. Gajewki, J. Oggonowski, and E. Roth, *Czas. Tech. M* **3**, 34 (1974).

76. A. A. Grigor'ev and co-workers, *Nauchno-Issled. Inst. Sint. Spirtov Org. Prod.* **6**, 65 (1974).

77. **Brit. Pat. 574,446** (Jan. 7, 1946), R. N. Lacey.

78. **Jpn. Kokai 72 15,810** (May 11, 1972), K. Takagi and K. Manabe (to Sumitomo Chemical Co. Ltd.).

79. **Jpn. Kokai 72 15,809** (May 11, 1972), K. Takagi and K. Manabe (to Sumitomo Chemical Co. Ltd.).

80. A. Pirvulescu, M. Panaitescu, and T. Boata, *Riv. Chim.* **12**, 377 (1961).

81. **Jpn. Kokai 72 15,808** (May 11, 1972), K. Tayagi and K. Manabe (to Sumitomo Chemical Co. Ltd.).

82. **U.S. Pat. 3,480,659** (Nov. 25, 1969), (to Shell Oil).

83. K. Schmitt, *Chem. Ind. (Dusseldorf)* **18**, 4, 204 (1966).

84. **U.S. Pat. 3,953,517** (Apr. 27, 1976), K. Schmitt and co-workers (to Veba-Chemie AG).

85. **Austrian Pat. 283,298** (July 27, 1970), (to Veba-Chemie AG).

86. **Brit. Pat. 227,977** (Apr. 15, 1971), (to Scholven Chemie).

87. **U.S. Pat. 3,574,763** (Apr. 13, 1971), J. Wollner and W. Neier (to Rheinpreussen).

88. **DE Pat. 1,238,453** (Nov. 9, 1967), H. Giehling (to Rheinpreussen).

89. **DE Pat. 1,193,931** (June 30, 1963), J. Wollner and F. Engelhardt (to Rheinpreussen).

90. *Chem. Ind. (Dusseldorf)*, 632 (Oct. 1981).

91. *Hydrocarbon Proc.*, 184 (Nov. 1977).

92. *Hydrocarbon Proc.*, 52 (Dec. 1985).

93. Y. Onoue and co-workers, *ChemTech.*, 36 (Jan. 1977).

94. Y. Watanabe and co-workers, *J. Catal.* **40**, 76 (1975).

95. Y. Watanabe and co-workers, *Bull. Chem. Soc. Jpn.* **47**, 12, 2922 (1974).

96. **Rus. Pat. 74006994** (Feb. 18, 1971), (to Tokoyama Soda).

97. **U.S. Pat. 3,946,079** (Nov. 23, 1976), (to Tokoyama Soda).

98. **Jpn Pat. 61 78,745** (Apr. 22, 1986), (to Sumitomo Chemical Co. Ltd.); K. Tanabe, *ChemTech.*, 628 (Oct. 1991).

99. **Eur. Pat. 271,182** (June 15, 1988), (to Sumitomo Chemical Co. Ltd.).

100. **Rus. Pat. 03 93,743** (Apr. 18, 1991), (to Sumitomo Chemical Co. Ltd.).

101. **Rus. Pat. 63 68,538** (Mar. 28, 1988), (to Mitsubishi Chem. Co. Ltd.).

102. **Brit. Pat. 1,226,029** (Mar. 24, 1971), (to British Petroleum Chemical Ltd.).

103. **Neth. Pat. Appl. 69,03635** (Mar. 11, 1969), (to Shell).

104. **Neth. Pat. Appl. 70,14622** (Oct. 5, 1976), (to Shell).

105. Voronov and co-workers, *USSR Klim. Provost (Moscow)* **8**, 477 (1986).

106. **U.S. Pat. 2,499,172** (Feb. 28, 1950), E. F. Smith (to Commercial Solvents).

107. **Brit. Pat. 1,015,003** (Dec. 31, 1965), (to Distillers).

108. **Rus. Pat. 46 2,009** (Jan. 19, 1971), (to Showa Denko).

109. **USSR Pat. 445,263** (Mar. 25, 1978), T. M. Khanonov and co-workers.

110. **Rus. Pat. 52 35,646** (Sept. 10, 1977), T. Kiyoura and co-workers (to Mitsui-Toatsu Chemical Co.).

111. *Chem. Mark. Rep.* (Aug. 20, 1990).

112. G. T. Austin, *Chem. Eng.*, 149 (June 24, 1974).

113. **U.S. Pat. 5,176,878** (Jan. 5, 1993), (to Teledyne Industries, Inc.).

114. *Hydrocarbon Proc.*, 211 (Nov. 1992).

115. S. K. Datta, *Indian Chem. J.* **9**, 8, 535 (1975).

116. E. C. Craven, *J. Appl. Chem. (London)* **13**, 71 (1963).

117. S. Kudo and co-workers, *Chem. Eng. Japan* **30**, 12, 1124 (1966).

118. **Brit. Pat. 1,527,033** (Oct. 4, 1978), C. Hawkins and B. Yeomans (to B.P. Chemicals Ltd.).

119. **U.S. Pat. 2,889,369** (June 6, 1959), P. Godet (to Usines de Melle).

120. **Ger. Pat. 1,052,970** (Mar. 19, 1959), K. Schmitt and J. Disteldorf (to Bergwerksgesellschaft Hibernia AG).

121. **Rus. Pat. Kokai 58 456** (Jan. 30, 1958), S. Yamada, S. Noguchi, and S. Tago (to Nippon Oil Co.).

122. J. Herscovici, T. Bota, and D. Sirenteanu, *Rev. Chim. (Bucharest)* **15**, 12, 736 (1964).

123. **U.S. Pat. 1,550,792** (Aug. 25, 1926), W. J. Edmonds.

124. **Brit. Pat. 504,337** (Apr. 24, 1939), S. H. McAllister and E. F. Bullard (to N. V. de Bataafsche Petroleum Maatschappi).

125. R. A. Garcia, J. V. Sinistera, and J. M. Marinas, *React. Kinet. Catal. Lett.* **18**, 33 (1981).

126. B. P. Ershov and V. L. Pridoragin, *J. Appl. Chem. (USSR)* **19**, 38 (1946).

127. **Ger. Pat. 1,052,970** (Nov. 27, 1950), W. Schlenk (to BASF).

128. I. Ichikizaki, N. Kawamura, and T. Hoshino, *Chem. High Polym.* **5**, 75 (1948).

129. **Jpn. Kokai 72 18,731** (Nov. 5, 1968), S. Hattori and co-workers (to Mitsubishi Chem. Ind. Co.).

130. **Jpn. Kokai 72 18,730** (Nov. 5, 1968), S. Hattori and co-workers (to Mitsubishi Chem. Ind. Co.).

131. J. Przondo, E. Bielous, and I. Franek, *Przem. Chem.* **59**, 436 (1980).

132. **Jpn. Pat. 73 52,715** (July 24, 1973), Y. Moriyasu and co-workers (to Mitsubishi Chem. Ind. Co.).

133. P. W. Sherwood, *Pet. Ref.* **33**, 12, 144 (1954).

134. **Jpn. Pat. 80 108,831** (Aug. 21, 1980), (to Mitsui Petrochemical Ind.).

135. H. Matyschok and S. Ropuszynski, *Chem. Stosov. Ser. A*, **12**, 283 (1986).

136. Z. N. Verkhovskaya and co-workers, *Khim. Prom.* **43**, 500 (1967).

137. **S.U. Pat. 963,983** (Oct. 7, 1982), A. P. Khordin and co-workers.

138. **Jpn. Pat. 58 59,935** (Apr. 9, 1983), (to Kyowa Yuka Co. Ltd.).

139. J. Przondo and co-workers, *Przem. Chem.* **51**, 8, 504 (1972).

140. **U.S. Pat. 1,714,378** (May 21, 1929), A. Korr and A. Weissenborn (to Winthrop Chem. Co.).

141. **Brit. Pat. 881,918** (Nov. 8, 1961), H. D. Burgess (to Union Carbide Corp.).

142. **Brit. Pat. 968,151** (Sept. 2, 1964), K. Hirai (to Nippon Sekiyu).

143. **H.U. Pat. 38,089** (Apr. 28, 1986), B. Jover and co-workers.

144. **Jpn. Pat. 61 05,040** (Jan. 10, 1986), (to Sumitomo Chemical Co. Ltd.).

145. **Jpn. Pat. 58 59,936** (Apr. 9, 1983), (to Kyowa Yuka Co. Ltd.).

146. E. E. Connolly, *J. Chem. Soc.*, 338 (1944).

147. **Jpn. Pat. 73 72,110** (Sept. 29, 1973), S. Kuwata and T. Uchida (to Mitsubishi Chem. Ind. Co. Ltd.).

148. **Jpn. Pat. 59,222,435** (Dec. 14, 1984), (to Mitsui Petrochemical Ind. Ltd.).

149. H. H. Szmant, *Organic Building Blocks of the Chemical Industry*, John Wiley & Sons, Inc., New York, 1989, p. 315.

150. **Brit. Pat. 990,434** (Apr. 28, 1965), M. W. Wirth (to British Hydrocarbon Chem. Ltd.).

151. **Jpn. Pat. 72 18,728** (Nov. 29, 1968), S. Kumano and A. Yasui (to Kuraray Co. Ltd.).

152. **DP Pat. 1,233,848** (1964), J. Wollner and F. Engelhardt (to Rheinpreussen).

153. **DE Pat. 3,702,483** (Aug. 11, 1988), W. Hoelderich and co-workers (to BASF AG).

154. **Eur. Pat. Appl. 209,785** (Jan. 28, 1987), W. Hoelderich and co-workers (to BASF AG).

155. **S.U. Pat. 791,732** (Dec. 30, 1980), S. I. Kryukov and co-workers.

156. **S.U. Pat. 144,167** (Feb. 6, 1962), E. G. Rozantsev.

157. *Soap. Cosmet. Chem. Spec.* **61**, 7, 85 (1985).

158. **U.S. Pat. 4,146,581** (Mar. 27, 1979), W. Fliege and co-workers (to BASF AG).

159. **U.S. Pat. 2,615,921** (Oct. 28, 1952), C. C. Chapman and C. F. Dougherty, Jr. (to Philips Petroleum).

160. **U.S. Pat. 3,758, 557** (Sept. 11, 1973), J. Hanotier and co-workers (to Labofina).

161. C. Dudley and G. Read, *Tetrahedron Lett.* **52**, 5273 (1972).

162. **Jpn. Pat. 79 95,816** (July 27, 1979), (to Mitsubishi Gas Chemical Co. Ltd.).

163. *Chem. Mark. Rep.*, 34 (Nov. 3, 1986).

164. **DE Pat. 2,625,541** (Dec. 22, 1977), A. Nissen and co-workers (to BASF AG).

165. **U.S. Pat. 4,754,074** (June 28, 1988), G. P. Hussmann (to Amoco).

166. **Jpn. Pat. 61,207,354** (Sept. 13, 1986), K. Matsuoka and K. Tagawa (to Daciel Chemical Industries Ltd.).

167. **DE Pat. 2,758,113** (July 5, 1979), H. Froehlich and co-workers (to BASF AG).

168. **DE Pat. 2,111,722** (Sept. 14, 1972), L. Schuster and L. Arnold (to BASF AG).

169. **U.S. Pat. 4,945,185** (July 31, 1990), D. R. Bryant and A. J. Papa (to Union Carbide Corp.).

170. **Brit. Pat. Appl. 2,202,165** (Sept. 21, 1988), E. Drent (to Shell Internationale Research Maatschappij BV).

171. **DE Pat. 2,445,193** (Apr. 4, 1976), (to BASF AG).

172. M. Hauser, *Chem. Rev.* **63**, 311 (1963).

173. A. Reiche, E. Schmitz, and E. Grunderman, *Chem. Ber.* **93**, 2443 (1960).

174. F. H. Stross, J. M. Monger, and H. de Finch, *J. Am. Chem. Soc.* **69**, 1627 (1947).

175. J. C. Aumiller and J. A. Whittle, *J. Org. Chem.* **41**, 2959 (1976).

176. D. S. Noyce and M. Evett, *J. Org. Chem.* **37**, 397 (1972).

177. J. Przondo and E. Bielous, *Prøgm. Chem.* **53**, 9, 541 (1974).

178. **U.S. Pat. 2,827,490** (Mar. 18, 1958), A. W. Martin (to Celanese).

179. **Rom. Pat. 44,576** (Dec. 13, 1966), J. Herscovici and co-workers (to Romanian Ministry of the Chemical Industry).

180. **U.S. Pat. 1,913,159** (Dec. 23, 1929), H. Guinot (to Usinert de Melle).

181. Y. K. Kim and J. D. Hatfield, *J. Chem. Eng. Data* **30**, 149 (1985).

182. **Brit. Pat. 1,116,037** (June 6, 1968), L. J. Sirois (to Esso Research and Engineering).

183. **U.S. Pat. 3,153,068** (Oct. 13, 1964), W. J. Porter and co-workers (to Esso Research and Engineering).

184. F. G. Klein and J. T. Banchero, *Ind. Eng. Chem.* **48**, 1278 (1956).

185. **Jpn. Pat. 45 41,566** (Sept. 18, 1970), Y. Mizutani (to Tokuyama Soda).

186. G. S. Salvapati and co-workers, *J. Mol. Catal.* **54**, 9 (1989).

187. S. Ito, *Industrial Toxicological Studies on Mesityl Oxide (Yokohama)* **20**, 3, 253 (1969).

188. E. C. Craven, *J. Appl. Chem.* **12**, 120 (1962).

189. **U.S. Pat. 4,005,145** (Jan. 25, 1977), E. Widmer (to Hoffmann-LaRoche, Inc.).

190. **U.S. Pat. 4,845,303** (July 4, 1989), H. Bellut (to HbIs).

191. J. N. Marx, *Org. Preparations Procedures Int.* **5**, 1, 45 (1973).

192. D. E. Loeber and co-workers, *J. Chem. Soc., C*, 404 (1971).

193. G. S. Salvapati and M. Janardanarao, *J. Sci. Ind. Res.* **42**, 5, 261 (1983).

194. H. Matyschok, *Wiad. Chem. (Pol)* **27**, 5, 311 (1973).

195. **U.S. Pat. 3,337,633** (Aug. 22, 1967), K. Schmitt and co-workers (to Hibernia Chemie GmbH).

196. **U.S. Pat. 3,337,423** (Aug. 22, 1967), K. Schmitt and co-workers (to Hibernia Chemie GmbH).

197. **U.S. Pat. 3,981,918** (Sept. 21, 1976), J. R. Walton and B. Yeomans (to B.P. Chemicals International, Ltd.).

198. **U.S. Pat. 4,059,632** (Nov. 22, 1977), C. Cane and B. Yeomans (to B.P. Chemicals International, Ltd.).

199. S.U. Pat. 1,074,859 (Feb. 23, 1984), (to Erevan).

200. **Jpn. Pat. 50/6470** (Mar. 14, 1975), (to Mitsubishi Chemical Ind. Co. Ltd.).

201. **Jpn. Pat. 47/34344** (Nov. 21, 1972), (to Mitsui Toatsu Chemicals Co. Ltd.).

202. **U.S. Pat. 2,399,976** (May 7, 1946), (to Shell).

203. **Brit. Pat. 733,650** (July 13, 1955), (to Societe Industrielle Des Derives de l'Acetylene).

204. **Ger. Pat. 145,096** (Nov. 19, 1980), (to VEB Leuna-Werke).

205. **U.S. Pat. 3,497, 558** (Feb 24, 1970), N. Gravino, G. Kohan, and I. Palmer (to Gulf Oil Canada).

206. **U.S. Pat. 2,373,510** (June 23, 1941), W. A. Bailey, Jr. (to Shell).

207. **U.S. Pat. 2,419,142** (June 29, 1944), V. N. Ipatieff (to Universal Oil Products Co.).

208. **U.S. Pat. 2,183,127** (Mar. 19, 1938), T. H. Vaughn (to Union Carbide Corp.).

209. **U.S. Pat. 4,165,339** (Aug. 21, 1979), W. T. Reichle (to Union Carbide Corp.).

210. **U.S. Pat. 4,535,187** (Aug. 13, 1985), A. J. Papa and co-workers (to Union Carbide Corp.).

211. **U.S. Pat. 4,476,324** (Oct. 9, 1984), W. T. Reichle (to Union Carbide Corp.).

212. **U.S. Pat. 4,970,191** (Nov. 13, 1990), to A. A. Schutz (to Aristech Chem. Corp.).

213. **U.S. Pat. 5,153,156** (Oct. 6, 1992), to A. A. Schutz and L. A. Cullo (to Aristech Chemical Corp.).

214. **Eur. Pat. 512 843 A1** (Nov. 11, 1992), Fukao and co-workers (to Sumitomo Chemical Co.).

215. E. C. Craven and W. R. Ward, *J. Appl. Chem.* **10**, 18 (1960).

216. H. G. Franck and co-workers, *Liebigs Ann. Chem.* **724**, 94 (1969).

217. A. Roger and co-workers, *Bull. Soc. Chim. Fr.* **6**, 2590 (1968).

218. E. Cytot, *Ann Chim.* **6**, 413 (1971).

219. **U.S. Pat. 2,419,051** (Dec. 5, 1944), S. A. Ballard (to Shell).

220. **U.S. Pat. 2,968,677** (Nov. 12, 1957), M. W. Fewlass (to Distillers Co. Ltd.).

221. **Jpn. Pat. 62/72644** (Apr. 3, 1987), (to Daciel Chemical Ind.).

222. **Brit. Pat. 832,124** (Apr. 6, 1960), E. C. Craven and M. W. Fewlass (to Distillers Co., Ltd.).

223. **U.S. Pat. 2,566,792** (May 25, 1948), H. Dannenberg (to Shell).

224. **U.S. Pat. 2,444,006** (Mar. 13, 1945), H. Dannenberg (to Shell).

225. **Brit. Pat. 2,029,415** (Mar. 19, 1980), H. C. Bailey and J. S. Bartlett (to B.P.).

226. **U.S. Pat. 4,434,301** (Feb. 28, 1984), A. J. Papa (to Union Carbide Corp.).

227. **U.S. Pat. 4,248,673** (Feb. 2, 1981), (to Kuhlmann Produits Chimiques).

228. **Jpn. Pat. 61/289055** (Dec. 19, 1986), (to Daciel Chemicals Ind.).

229. Acetone Chemistry, *Technical Information Bulletin*, Veba-Chemie AG, Gelsenkirchen-Buer, Germany, 1978.

230. *Chem. Week* **21** (Sept. 30, 1992).

231. **Brit. Pat. 1,047,920** (Nov. 9, 1966), (to Schloven Chemie).

232. **U.S. Pat. 5,179,221** (Jan. 12, 1992), M. Koji and co-workers (Nippon Kasei Chemical Co.).

233. **U.S. Pat 5,091,554** (Feb. 25, 1992), K. Huthmacher and H. Schmitt, (to Degussa AG).

234. **U.S. Pat. 5,011,968** (Apr. 30, 1991), J. C. Thunberg and W. B. Begonis (to W. R. Grace Co.).

235. **U.S. Pat. 5,142,090** (Aug. 25, 1992), E. Pontoglio and S. Parodi (Caffaro SpA).

236. **Brit. Pat. 972,010** (Oct. 7, 1964), (to Hibernia-Chemie).

237. **Eur. Pat. 449,089** (Oct. 10, 1991), W. Harder and co-workers (to BASF AG).

238. **DE Pat. 1,202,785** (Oct. 14, 1965), K. Schmitt and co-workers (to Schloven Chemie).

239. *Chem. Week* **8** (May 20, 1992).

240. *Chem. Mark. Rep.* (Aug. 3, 1992).

241. IPDI *Technical Information Bulletin*, Veba-Chemie AG, Gelsenkirchen-Buer, Germany, 1975.

242. Ref. 148, p. 422.

243. **U.S. Pat. 3,446,850** (May 27, 1969), D. P. Cotrupe and co-workers (to ESSO Research and Engineering).

244. **U.S. Pat. 3,397,129** (Aug. 13, 1968), William and co-workers (to ESSO Research and Engineering).

245. **U.S. Pat. 3,361,822** (Jan. 2, 1968), K. Schmitt and co-workers (to Schloven-Chemie AG).

246. **S.U. Pat. 614,118** (June 6, 1978), G. D. Ballova, E. I. Egorova, and O. N. Romantsova.

247. **U.S. Pat. 3,894,848** (July 15, 1975), H. L. Kleinman and J. J. Martinat (to Technology Transfer, Ltd.).

248. **Jpn. Pat. 54 081240** (June 28, 1979), (to Kyowa Fermentation).

249. **U.S. Pat. 3,663,597** (May 16, 1972), D. Flitter (to American Home Products Co.).

250. U.S. 3,673,239 (June 27, 1972), H. Takahashi (to Daciel Ltd).

251. **Eur. Pat. 80,759** (June 8, 1983), A. J. C. Seters and co-workers (to Shell).

252. P. S. Sai Prasad and co-workers *Appl. Catal. A: General* **83**, 141 (1992).

253. **U.S. Pat. 3,816,546** (June 11, 1974), R. W. Rieve (to Atlantic Richfield Co.).

254. **U.S. Pat. 4,086,282** (Apr. 25, 1978), F. Wattimena (to Shell Oil).

255. **U.S. Pat. 4,188,497** (Feb. 12, 1980), Huizer and co-workers (to Shell Oil).

256. **Brit. Appl. 2,075,008** (Mar. 4, 1981), Kohler and co-workers (to Rutgerswerke).

257. H. G. Franck and J. W. Standelhofer, *Industrial Aromatic Chemistry*, Springer-Verlag, New York, 1988, p. 172.

258. Ref. 148, p. 512.

259. **Brit. Pat. 1,448,976** (Sept. 8, 1976), (to Hoffmann-LaRoche).

260. **U.S. Pat. 3,944,620** (Mar. 16, 1976), J. J. Becker, U. P. Hochstrasser, and W. Skorianetz (to Firmenich SA).

261. **U.S. Pat. 4,970,347** (Nov. 13, 1990), H. Bellut (to Huls AG).

262. **U.S. Pat. 4,026,948** (May 31, 1977), J. J. Becker and co-workers (to Firmenich S.A.).

263. **U.S. Pat. 4,046,813** (Sept. 6, 1977), W. Brenner (to Hoffmann-LaRoche).

264. National Toxicology Program Technical Report Series, NPT-TR-29,186, Research Triangle Park, N.C.

265. M. Sittig, *Handbook of Toxic and Hazardous Chemicals*, Noyes Publications, Park Ridge, N.J., 1981.

266. C. E. Schildknecht, *Vinyl and Related Polymers*, John Wiley & Sons, Inc., New York, 1952.

267. **U.S. Pat. 3,544,634** (Dec. 1, 1970), H. Muller, H. Koehl, and H. Pommer (to BASF AG).

268. **Jpn. Pat. 4,828,211** (1971), (to Mitsubishi).

269. *Kblim. Promst (Moscow)* **10**, 727 (1969).

270. **U.S. Pat. 3,364,264** (Jan. 16, 1968), H. F. Harley and R. K. Grasselli (to Standard Oil Co.).

271. **U.S. Pat. 1,967,225** (July 24, 1934), R. F. Conaway (to Du Pont).

272. **U.S. Pat. 2,267,829** (Dec. 30, 1941), H. Lange and O. Horn (to I.G. Farben).

273. *Code of Federal Regulations*, 29 CFR 1910.119, OSHA, Washington, D.C., May 26, 1992; *Chem. Eng.*, 108 (Jan. 1993).

274. **U.S. Pat. 3,422,148** (Jan. 14, 1969), J. Wollner and co-workers (to Rheinpreussen).

275. E.P. Pat. Appl. 348,793 (Jan. 3, 1990), W. Hoelderich and co-workers (to BASF AG).

276. **Eur. Pat. Appl. 296,488** (Dec. 28, 1988), W. Hoelderich and co-workers (to BASF AG).

277. **DE Pat. 3,632,530** (Apr. 7, 1988), W. Hoelderich and co-workers (to BASF AG).

278. **U.S. Pat. 2,829,168** (Apr. 1, 1958), F. J. Lowes (to Dow Chemical Co.).

279. E. MacMahon and co-workers, *J. Am. Chem. Soc.* **70**, 2971 (1948).

280. **Eur. Pat. Appl. 350,320** (Jan. 10, 1990), J. J. Gyuran (to Pharmaglobe Labs Ltd.).

281. *Chem. Mark. Rep.* (Feb. 15, 1988) and (Mar. 28, 1988).

282. J. Buckingham, ed., *Dictionary of Organic Compounds*, 5th ed., Chapman & Hall, New York, 1982.

283. *Catalog of Fine Chemicals*, Aldrich Chemical Co., Milwaukee, Wis., 1992–1993.

284. Y. Ogata and M. Yamashita, *Tetrahedron* **27**, 3395 (1971).

285. M. Stoll and co-workers, *Helv. Chim. Acta.* **50**, 628 (1967).

286. K. Kulka, *Am. Perfum.* **76**, 11, 23 (1961).

287. Neth. Pat. 63,040 (May 16, 1949), (to N. V. Bataafsche Petroleum Maatschappij).

288. S. B. Bowlus and J. A. Katzenellenbogen, *J. Org. Chem.* **39**, 3309 (1974).

289. E. J. Fairchild, ed., *Registry of Toxic Effects of Chemical Substances*, Vol. II, National Institute for Occupational Safety and Health (NIOSH), U.S. Dept. of Health Education and Welfare, Cincinnati, Ohio, Sept. 1977.

290. **U.S. Pat. 2,455,631** (Dec. 7, 1948), O. Weinkauff (to Monsanto Chemical Co.).

291. **D.P. Pat. 2,831,229** (1978), Saner and co-workers (to BASF).

292. **U.S. Pat. 2,062,263** (Nov. 24, 1937), E. Eberhardt (to I.G. Farbenindustrie).

293. **U.S. Pat. 1,899,094** (Feb. 28, 1933), A. J. Kluyver and M. A. Scheffer (to Thomas H. Verhane).

294. **U.S. Pat. 2,799,707** (July 16, 1956), O. C. Dermer (to Dow Chemical Co.).

295. H. Soumalainen and L. Jannes, *Nature* **157**, 335 (1946).

296. **U.S. Pat. 5,075,226** (Dec. 24, 1991), K. Tsutomoto and co-workers (to Meiji Milk Products Co.).

297. **U.S. Pat. 4,867,992** (Sept. 19, 1989), G. Bertkam and co-workers (to General Foods Corp.).

298. **U.S. Pat. 2,377,749** (June 5, 1945), C. A. Bordner (to E. I. du Pont de Nemours & Co., Inc.).

299. **U.S. Pat. 2,658,920** (Nov. 10, 1953), D. X. Klein and co-workers (to Heyden Chemical Corp.).

300. Ref. 258, p. 257.

301. *Jpn. Chem. Week* **3** (Aug. 19, 1976).

302. F. M. Snyder and F. A. Morton, *J. Econ. Entomol.* **39**, 385 (1946).

303. K. Kulka, *Am. Perfum.* **78**, 9, 27 (1963).

304. W. M. Muir, P. D. Ritchie, and D. J. Lyman, *J. Org. Chem.* **31**, 3790 (1966).

305. D. C. Nonhebel and H. D. Murdoch, *J. Chem. Soc.*, 2153 (1962).

306. C. Weygand and H. Baumgartel, *Chem. Ber.* **62B**, 574 (1929).

307. **U.S. Pat. 2,867,653** (Jan. 6, 1959), E. Enk and F. Buttner (to Wacker-Chemie).

308. **U.S. Pat. 2,511,423** (1950), F. G. Young (to Carbide and Carbon Chem. Co.).

309. **U.S. Pat. 2,395,800** (Mar. 5, 1946), A. B. Boese, Jr., and F. G. Young, Jr. (to Union Carbide Corp.).

310. **U.S. Pat. 2,787,642** (Apr. 2, 1957), E. Enk and F. Buttner (to Wacker-Chemie).

311. **U.S. Pat. 3,794,686** (Feb. 26, 1974), H. Spes and G. Kunstle (to Wacker-Chemie).

312. L. Nowakowski and L. Glowienkowska, *Chemik (Pol.)* **36**, 1, 279 (1983).

313. K. K. Georgieff, *Ind. Eng. Chem.* **49**, 1067 (1957).

314. **Brit. Pat. 681,696** (1952), R. Decker and H. Holz.

315. **Brit. Pat. 741,563** (1955), (to Wacker-Chemie).

316. **Czech. Pat. 91,577** (Sept. 15, 1959), A. Novacek and B. Vondracek.

317. S.U. 235,005 (Jan. 16, 1969), Yashunskii and co-workers.

318. **U.S. Pat. 2,395,012** (Feb. 19, 1946), W. H. Reeder and G. A. Lescisin (to Union Carbide Corp.).

319. **U.S. Pat. 2,432,499** (Dec. 16, 1947), A. B. Boese, Jr. (to Union Carbide Corp.).

320. **Brit. Pat. 1,213,255** (Nov. 25, 1970), (to Wacker-Chemie).

321. **Brit. Pat. 1,213,255** (Nov. 25, 1970), (to Wacker-Chemie).

322. **U.S. Pat. 2,737,528** (Mar. 6, 1956), C. F. Hunt (to Shawingan Chemicals Ltd.).

323. **U.S. Pat. 2,834,811** (May 13, 1958), K. K. Georgieff (to Shawingan Chemicals Ltd.).

324. **Brit. Pat. 934,375** (Aug. 21, 1963), (to Shell).

325. *Material Safety Data Sheet*, Union Carbide Corp. Danbury, Conn., Jan. 1, 1992.

326. M. S. Kharasch, H. C. McBay, and W. H. Urry, *J. Am. Chem. Soc.* **70**, 1269 (1948).

327. Y. Koybayashi, T. Taguchi, and T. Tokuno, *Tetrahedron Lett.* **42**, 3741 (1977).

328. Y. Ito, T. Konoike, and T. Saegusa, *J. Am. Chem. Soc.* **99**, 1487 (1977).

329. **DE Pat. 2,437,219** (1974), H. Stetter and H. Kuhlmann (to Bayer AG).

330. C. S. Subramaniam and co-workers, *J. Chem. Soc. Perkin Trans. I.*, 2346 (1979).

331. J. E. McMurry and T. E. Glass, *Tetrahedron Lett.* **27**, 2575 (1971).

332. **U.S. Pat. 2,052,652** (1936), Perkins and Toussaint (to Union Carbide Corp.).

333. F. B. Wells and C. F. H. Allen, *Org. Syn., Coll. Vol II*, 219 (1943).

334. **DE Pat. 2,831,229** (Jan. 31, 1980), W. Sauer and co-workers (to BASF AG).

335. **Jpn. Pat. 72 11,411** (Mar. 7, 1968), T. Konaka and co-workers (Shionogi and Co., Ltd.).

336. **Brit. Pat. 1,460,544** (Jan. 6, 1977), R. H. Lasco (to Diamond Shamrock Corp.).

337. D. Lloyd, *Alicyclic Compounds*, American Publishing Co., Inc., New York, 1963, p. 68.

338. S. Coffey, ed., *Chemistry of Carbon Compounds*, 2nd ed., Vol. 1, Part B, Elsevier Publishing Co., New York, 1968, p. 409.

339. G. H. Whitham, *Alicyclic Chemistry*, Oldbourne Press, London, 1963, p. 7.

340. A. Gero, *J. Org. Chem.* **26**, 3156 (1961).

341. **U.S. Pat. 3,968,162** (July 6, 1976), P. E. Bost and G. Lartigan (to Rh  ne Poulenc SA).

342. H. F. Keag, H. S. McCullough, and H. J. Sanders, *Ind. Eng. Chem.* **45**, 2 (Jan. 1953).

343. **U.S. Pat. 2,797,245** (June 25, 1957), V. Hasnsel (to UOP Inc.).

344. *Eur. Chem.*, 250 (June 7, 1989).

345. K. Bauer and D. Garbe, *Common Fragrance and Flavor Materials*, VCH-Publishers, New York, 1985, p. 73.

346. W. T. Read, *J. Am. Chem. Soc.* **44**, 1751 (1922).

347. **U.S. Pat. 4,177,097** (Oct. 23, 1979), C. A. Smith and L. F. Theiling (to Union Carbide Corp.).

348. *Chem. Mark. Rep.* (May 5, 1980).

349. *Specialty Chemical Handbook*, 2nd ed., Japan Chemical Week (The Chemical Daily Co., Ltd.), Tokyo, Mar. 1987.

350. **U.S. Pat. 3,642,906** (1969), S. Kuhn (to Universal Oil Products Co.).

351. **U.S. Pat. 2,528,789** (Nov. 7, 1950), P. D. Caesar and A. M. Sachanen (to Mobil Corp.).

352. *Chim. Actual* **302**, 6 (Nov. 12, 1989).

353. *Am. Paint Coatings J.* **74**, 50, 16 (May 21, 1990).

354. *Chem. Mark. Rep.* **11** (Sept. 18, 1989).

John Braithwaite
Union Carbide Chemicals and Plastics Company, Inc.

KINETIC MEASUREMENTS

Kinetic measurements are studies of the rates at which chemical reactions occur. Generally, these studies involve preparing a chemical system using reagent concentrations different from the equilibrium values and then monitoring the concentration changes as the system approaches equilibrium, although other, less direct strategies are sometimes exploited. Chemical kinetic data are used in materials sciences, biochemistry and molecular biology, earth and atmospheric science, and many branches of engineering. Related concepts appear in nuclear physics, but presuppositions and methods are different there.

Kinetic information is acquired for two different purposes. First, data are needed for specific modeling applications that extend beyond chemical theory (see [ATMOSPHERIC MODELING](#); [ENGINEERING, CHEMICAL DATA CORRELATION](#)). These are essential in the design of practical industrial processes, and are also used to interpret natural phenomena such as the observed depletion of stratospheric ozone (qv). Compilations of measured rate constants are published in the United States by the National Institute of Standards and Technology (NIST), similar to data on equilibrium thermodynamic properties (qv) published by the same organization and its predecessor, the National Bureau of Standards. Second, kinetic measurements are undertaken to elucidate basic mechanisms of chemical change, simply to understand the physical world. The ultimate goal is control of reactions, but the immediate significance lies in the patterns of kinetic behavior and the interpretation in terms of microscopic models.

Explaining chemical change by postulating mechanisms and testing against measurements is expected to continue to dominate most of chemistry for the foreseeable future. For the specific goal of providing a detailed understanding of simple reactions, however, traditional kinetics is being challenged by theoretical and experimental methods that focus directly on the behavior of individual atoms. Proponents of this newer viewpoint do not refer to themselves as kineticists. No name is fully established, but this discipline is often referred to as reaction dynamics.

Macroscopic Behavior and the Rate Law

Chemical Equations Chemical changes are discussed with the aid of the equations used to treat equilibrium, ie, the reaction of reactants *A*, *B*, *C*, and so on, to produce products *P*, *Q*, and so forth,

$$aA + bB + cC + \dots \rightleftharpoons pP + qQ + \dots$$

(1)

The components *A*, *B*, *P*, *Q*, ... may be atoms, molecules, or ions. Kinetic rates are sensitive to a host of factors that must be specified or inferred, such as temperature, pressure, and presence of inert solvent or active catalyst. Most often, a kinetic change is written so that there is an initial excess of reactants which decrease over time.

The essential information implied by the chemical equation is the stoichiometry at the macroscopic level, ie, if *a* moles of *A* react, then *b* moles of *B* do also; *p* moles of *P* are formed, etc. No inference should be made about behavior at the microscopic or atomic level, ie, there is no implication that *p* molecules of *P* appear simultaneously. There may or may not be intermediates that appear and disappear in the course of the reaction.

Traditional chemical kinetics uses notation that is most satisfactory in two cases: all components are gases with or without an inert buffer gas, or all components are solutes in a liquid solvent. In these cases, molar concentrations represented by brackets, are defined, which are either constant throughout the system or at least locally meaningful. The reaction quotient *Z* is defined as

$$Z = \frac{[P]^p [Q]^q \dots}{[A]^a [B]^b [C]^c \dots}$$

(2)

The reaction quotient may be measured, at least in principle, for the reacting system at any time. If *Z* is observed not to change, the system is at equilibrium, or trapped in a metastable state that serves as a local equilibrium. In informal work, a time-independent *Z* is identified directly with the equilibrium constant *K_c*.

A kinetic study typically prepares some initial *Z_i* not equal to *K_c* and describes the subsequent evolution of each of the concentrations. A basic assumption is that each component evolves according to some differential equation where *t* represents time.

$$d[A] / dt = f([A], [B], \dots, [P], \dots, \text{other conditions})$$

(3)

Establishing precisely the conditions necessary to justify the transition from a microscopic, quantum description to such a macroscopic differential equation is an interesting question in theoretical physics. It has been treated in considerable detail (1) but rarely troubles practicing kineticists.

Other conventions for treating equilibrium exist and, in fact, a rigorous thermodynamic treatment differs in important ways. For reactions in the gas phase, partial pressures of components are related to molar concentrations, and an equilibrium constant *K_p*, expressed directly in terms of pressures, is convenient. If the ideal gas law applies, the partial pressure is related to the molar concentration by a factor of *RT*, the gas constant times temperature, raised to the power of the reaction coefficients.

$$K_p = K_c (RT)^{(p+q+\dots) - (a+b+c+\dots)}$$

(4)

Only those components which are gases contribute to powers of *RT*. More fundamentally, the equilibrium constant should be defined only after standard states are specified, the factors in the equilibrium constant should be ratios of concentrations or pressures to those of the standard states, the equilibrium constant should be dimensionless, and all references to pressures or concentrations should really be references to fugacities or activities. For reactions involving moderately concentrated ionic species (>1 mM) or moderately large molecules at high pressures (~1–10 MPa), the activity and fugacity corrections become important; in those instances, kineticists do use the proper relations. In some other situations, eg, reactions on a surface, measures of chemical activity must be introduced. Such cases may often be treated by straightforward modifications of the basic approach covered herein.

The focus herein is a survey of contemporary experimental approaches to determining the form of equation 3 and quantifying the parameters. In general, the differential equation could be very complicated, eg, the concentrations may be functions of spatial coordinates as well as time. Experimental measurements are arranged to ensure that simplified equations apply.

The Well-Stirred Mixture A key assumption of most kinetic measurements is that of a well-mixed solution of reactants. Then any component can be characterized by a single time-dependent concentration, applicable to the entire system. This assumption can usually be justified for gases or low viscosity liquids in volumes of laboratory dimensions over times not too short; typically >1 ns for low viscosity solvents such as water. It greatly simplifies interpretation of experiments and extraction of rate constants. Mixing may be accomplished by relying on passive diffusion of components or by actively stirring the mixture (see [MIXING AND BLENDING](#)). The assumption cannot, however, be maintained in all circumstances. Engineers must combine chemical kinetic data with the description of mass transport to design practical reactors (see [REACTOR TECHNOLOGY](#)). They may also use heterogeneous catalysts and need to treat surfaces explicitly. Geologists considering the formation of minerals consider very slow diffusion of reactants.

Biologists are quite sure that reagents are not homogeneously mixed in living systems. Chemical physicists striving to understand the ultrafast, primary processes of chemical change know that uniform concentration is too crude an approximation for their purposes. Nonetheless, the assumption of a well-stirred mixture is so pervasive that kineticists rarely point it out in their reports.

The Rate Law The goal of chemical kinetic measurements for well-stirred mixtures is to validate a particular functional form of the rate law and determine numerical values for one or more rate constants that appear in the rate law. Frequently, reactant concentrations appear raised to some power. Equation 5 is a rate law, or rate equation, in differential form.

$$d[P] \over dt = k[A]^x[B]^y \dots [P]^w \dots \tag{5}$$

The exponents describe the order of the reaction. It is said to be *x*-order in *[A]*, *y*-order in *[B]*, and (*x* + *y* + . . . + *w* + . . .)-order overall. The exponents may be positive, negative, or zero. A reaction zero-order with respect to a component proceeds at the same rate regardless of the amount of that component present. For example, if a reaction goes to completion, with equilibrium lying far to the right, then the rate law may well be independent of the amount of any products, and may be said to be zero-order in all products. Most likely, the products would simply be omitted from the rate law, as it would be written. There are other ways that a reaction may become zero-order in a particular component, at least over some limited range of concentrations. The exponents are often integers, but may be simple fractions like 3/2 or 1/3. Some rate laws do not allow specifying reaction orders over a wide concentration range, but may show different apparent reaction orders in limiting regimes, for example,

$$d[P] \over dt = k_1[A]^x (k_2 + [B]^y) \tag{6}$$

The rate law draws attention to the role of component concentrations. All other influences are lumped into coefficients *k*₁, called reaction rate constants. The *k*₁ are not supposed to change as concentrations change during the course of the reaction. Although *k*₁ are referred to as rate constants, they change with temperature, solvent, and other reaction conditions, even if the form of the rate law remains the same.

Additionally, the following points are important: (1) there is no guaranteed relation between the exponents *x*, *y*, *w*, . . . and the stoichiometric coefficients *a*, *b*, *p*, Only in special circumstances are there relationships. (2) There is no guarantee that the same rate law applies over any range of circumstances beyond what is demonstrated. It is common for different rate laws to be needed in different regimes of temperature or concentration, even when the same stoichiometric chemical equation applies. The effect of a catalyst may be either an increase in some rate constant or a change of the form of the rate law itself. (3) Reactions may depend on the strength of surrounding electromagnetic fields, such as visible light. (4) A rate law may be written for each component. Relations among these are determined by the stoichiometric coefficients

$$(1/p) (d[P] \over dt) = (1/q) (d[Q] \over dt) = \dots (1/a) (d[A] \over dt) = \dots \tag{7}$$

Results may be reported for any component. The functional form of the rate law and the exponents *x*, *y*, *w*, . . . are not affected by such an arbitrary choice. The rate constants, however, may change in numerical value. Similarly, the stoichiometric chemical equation may be written in alternative but equivalent forms. This also affects, at most, the numerical value of rate constants. Consequently, one must know the chemical equation assumed before using any rate constant.

Integrated Forms of Rate Laws Given a rate law, valid over the time range of interest, and a set of initial conditions, integration of the differential equation yields concentrations expressed as functions of time. In the past, considerable effort was devoted to cataloging integral solutions and training students in their interpretation. Because computers are able to perform numerical integrations very quickly and display time courses in graphical output, there is an increasing tendency to work with the differential forms and avoid writing analytical solutions for the integrated forms, except for simple cases.

Experimental Verification of a Rate Law

It is possible to prepare a system having an initial concentration for each component, and then measure a finite, but small, change in the concentration of one component, Δ*[A]* for example, over a known interval of time Δ*t*. The experimental velocity Δ*[A]* / Δ*t* and the concentrations can be substituted into a proposed rate law, along with postulated values for the exponents *x*, *y*, . . . to determine an observed rate constant *k*_{obs}, although the rate law may involve more than one *k*_{obs}. If this process is repeated for a reasonable range of concentrations, and the postulated rate law having the same exponents always yields the same *k*_{obs}, then it is asserted that the rate law has been verified and the rate constant has been determined, within some precision, and is valid for those concentration ranges. This approach is a reasonable strategy for an initial survey of a totally unknown system. Moreover, it avoids the need to know the endpoint of the reaction. It is, however, tedious and gives imprecise values for *k*_{obs}. It is also wasteful, in that it extracts very little data from each set of initial conditions. More often, the integrated form of the rate law is fit to multiple concentration measurements recorded at different times following each set of initial conditions.

Flooding and Pseudo-First-Order Conditions For an example, consider a reaction that is independent of product concentrations and has three reagents. If a large excess of *[B]* and *[C]* are used, and the disappearance of a lesser amount of *[A]* is measured, such flooding of the system with all components but *A* permits the rate law to be integrated with the assumption that all concentrations are constant except *[A]*. Consequently, simple expressions are derived for the time variation of *[A]*. Under flooding conditions and using equation 8, if *x* happens to be 1, the time-dependent concentration

$$d[A] \over dt = k_3[A]^x[B_t]^y[C_t]^z \tag{8}$$

of *A* exhibits an exponential decrease from its initial value *[A*_{*i*}] to its final equilibrium value, or endpoint, *[A*_∞]:

$$[A(t)] = [A_\infty] + ([A_i] - [A_\infty]) \exp(-k_{\text{obs}}t) \tag{9}$$

The conditions chosen make the reaction appear to be first-order overall, although the reaction is really not first-order overall, unless *y* and *z* happen to be zero. If a simple exponential is actually observed over a reasonable extent (at least 90–95%) of decay the assumptions are considered validated and *k*_{obs} is obtained with good precision. The pseudo-first-order rate constant *k*_{obs} is related to the *k* in the originally postulated rate law by

$$k_{\text{obs}} = k_3[B_t]^y[C_t]^z \tag{10}$$

If the same measurement is repeated for different *[B]* it should be possible to extract *y* by plotting log *k*_{obs} vs log *[B]*. This should be a straight line with slope *y*. In a similar manner, *z* can be obtained by varying *[C]*. At the same time the assumption that *x* equals 1 is confirmed. Ideally, a variety of permutations should be tested. Even if *x* is not 1, and the integrated rate equation is not a simple exponential, a useful simplification still results from flooding all components except one.

The endpoint value for any changing concentration, such as *[A*_∞], sometimes referred to as the infinity point, is extremely important in the data analysis, particularly when the order of the reaction is not certain. The obvious way to determine it, ie, by allowing the reaction to proceed for a long time, is not always reliable. It is possible for secondary reactions to interfere. It may sometimes be better to calculate the endpoint from a knowledge of the

equilibrium properties of the system.

General Considerations in Experimental Design It is convenient to be able to make measurements for the integrated rate equation over a time range of 10^4 using the same apparatus. There is some minimum time required to start the reaction and make the first concentration measurement. It is desirable, although not always necessary, that this instrument response time be 10 times faster than any reaction half-life, that is, the time needed for a twofold change in whatever concentration is being monitored. Then it is necessary to be able to continue to monitor the concentration for several half-lives. In addition, kinetic measurements require determining how a reaction rate changes as a function of reaction conditions. Finally, it is common to find that a reaction does not take place in a single step, but rather has a complicated time evolution. It is desirable to measure as much of the total behavior as possible in the same experiment. Consequently, in order to have a broadly useful, general-purpose instrument, a dynamic range of at least 10^4 is desirable.

Kinetic studies have benefited immensely from microcomputers. Whereas dedicated software is often necessary for interfacing to specific instruments, data analysis can be carried out using readily available software materials capable of producing high quality graphical output. Most recently, it has become common to measure concentrations in some way that produces digital data that is entered automatically into the computer (see [COMPUTER TECHNOLOGY](#)).

The Initial Conditions One of two very different strategies are used in kinetic measurements to produce the initial, nonequilibrium concentrations of reactants. Either the separate reagents are mixed or a system previously at equilibrium is perturbed. Each of these basic strategies has many variations.

Manual Mixing Methods The most obvious procedure for measuring reaction rates is simply to mix reactants together and then monitor either the disappearance of a reactant or appearance of a product. For reactions that are slow enough, mixing can be as simple as pouring liquid solutions into a flask or admitting gases into a reaction bulb. Aliquots may be removed from the mixture at selected time intervals, quenched, and the concentrations measured utilizing the same instrumentation used to measure static concentrations. The reaction vessel needs no special properties and may be inserted into a well-thermostated temperature bath. Quenching is done by quickly plunging the aliquot into a low temperature bath or by diluting it with excess solvent. Concentrations may be determined by volumetric titration, a classical method that is usually precise and accurate, or by a host of other methods (see [ANALYTICAL METHODS, SURVEY](#)). Quenched aliquots are also mandated when components must be separated by, for example, chromatography (qv), before being quantified. For gas-phase kinetics, measurement of total or partial pressures is a common strategy (see [PRESSURE MEASUREMENT](#)).

Manual mixing is applicable to a very narrow time range. A few minutes to several days is only a factor of 10^3 . It is desirable, therefore, to reduce the mixing time as much as possible. It is easy to mix low viscosity liquids or gases in less than a second by placing reactants in separate syringes, or the equivalent, and expelling them rapidly. With some attention to details of the flow through a mixer, times less than a millisecond can be achieved, multiplying the dynamic range of kinetic measurements by a factor of 10^5 , if the concentration analyses can be speeded up by a corresponding factor. The most dramatic improvements come from measuring concentrations directly in the reacting volume. However, it is sometimes necessary to make remote analyses. Although automation of the extractions from one reaction vessel is possible, it is easier to mix only small volumes of reactants at a time, quench the entire reaction volume after some delay, and repeat the process for a variety of delay times. Instrumentation is available commercially. All this, however, is still tedious, especially under exotic conditions, such as high or low temperature or pressure.

It is convenient to measure concentrations *in situ*. This requires an appropriately selective procedure. Spectrophotometry in the visible or uv is most common, although ir analysis may also be used and is becoming increasingly popular, because of its better selectivity (see [INFRARED TECHNOLOGY AND RAMAN SPECTROSCOPY](#); [SPECTROSCOPY](#)). Other regions of the spectrum are also used in special circumstances, for example, nuclear magnetic resonance (nmr) (see [MAGNETIC SPIN RESONANCE](#)). Conductivity measurements, which can be useful when ions are involved, are easy to automate, as are potentiometric and amperometric measurements (see [ELECTROANALYTICAL TECHNIQUES](#)). For gas-phase kinetics, the same spectroscopic methods are useful; in addition, a large variety of techniques depending on mass spectrometry (qv) are very attractive.

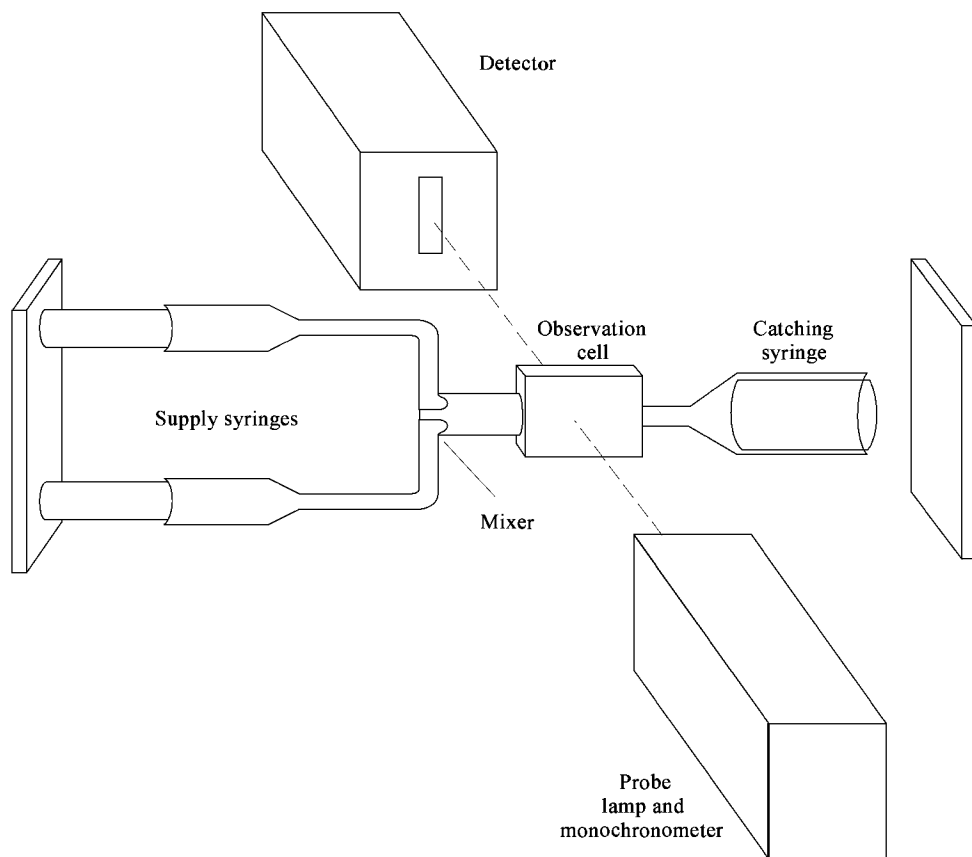


Fig. 1. Schematic of apparatus for stopped-flow measurements.

Stopped-Flow Mixing. Instruments that combine fast-acting syringes and good mixers with automated measurements of concentrations *in situ* are now readily available from several commercial sources. These are called stopped-flow apparatuses. Figure 1 shows a schematic diagram of the essential elements. Reactants are placed in separate syringes. The plungers are pushed forward very quickly, usually by means of high pressure air discharge. The two reagents are combined in the mixer and then flow into an observation cell, in which the concentration of a reactant or product is measured. A catching syringe is added to recover the sample and provide a controllable means of halting the high speed flow. The probe that measures concentrations as a function of time after mixing is almost always a uv or visible absorption spectrophotometer, although photofluorescence has become an option routinely available in commercial instruments. This latter measurement is actually a fluorescence-detected absorption measurement; the fluorescence is directly proportional to the amount of light absorbed. Conductivity measurement is also fairly common. Because the apparatus is quite compact, the entire

device can be thermostated easily over a range of at least 50°C around ambient, and with effort can be adapted for much more extreme conditions. Even devices operating at pressures up to a few hundred MPa are commercially available. Instruments often have three syringes, so that two reactants can be combined first and then a third component added after some delay.

Stopped-flow methods are a simple extension of classical benchtop methods, and have the very important advantage of using the minimum amount of reagents. This is especially desirable for biochemical investigations. The key to their invention was the development of electronic methods for measuring and recording concentrations in real time. In first generation instruments, the signal proportional to concentration was displayed on an oscilloscope and photographed. The photograph was subsequently analyzed. Relatively few points of limited precision could be extracted, and signal averaging was tedious. Contemporary designs use analogue-to-digital conversion of the electrical signal and direct transfer to a computer. Modern digitizers can capture thousands of sequential data points, easily attaining a dynamic range of 10⁴ in one measurement, taking less than a minute. Instruments usually include a two-way valve for each of the reactant syringes, so that these may be refilled easily to allow signal-averaging of five or ten measurements. Stopped-flow instruments are usually oriented toward liquid solution studies, and have become the standard kinetic procedure for biochemistry and bench-scale organic and inorganic solution chemistry, for reactions occurring over times longer than about 1 ms.

Flow Mixing. Before developments in fast electronics made stopped-flow measurements commonplace, continuous flow mixing was introduced for fast kinetic studies. In this procedure, reactants injected continuously into a flowing stream mix quickly to achieve the well-stirred condition. The reaction then proceeds as the mixture moves along a tube. The concentration of products grows with time, and therefore with position along the flow tube. Because the delay time after mixing is encoded into spatial position, the concentrations at one or more observation points remain constant and can be measured using slow analytic instruments, as the reacting mixture moves rapidly past. There is some difficulty achieving adequate dynamic range whether concentrations are measured at different points along the flow tube or the flow velocity is varied. Providing good temperature regulation for flow mixing can be unwieldy; and working at high pressures is certainly a challenge. The biggest intrinsic problem, however, is that flow mixing requires large volumes of reactants. It is still useful in special circumstances; and it is essential in chemical engineering in the design of industrial reactors, where not just the chemical kinetics but also the mass transport and mixing are part of the study.

Measurement Strategies Based on Perturbations

At times much less than a millisecond, it becomes impossible to achieve good mixing for two reasons: (1) it is increasingly difficult to move a liquid or gas sample even a small distance through a mixer, and (2) simultaneously, convective mixing needs to be ever more thorough in order to reduce the time needed for final diffusive mixing on the microscopic scale, which depends on the square of the distance over which diffusion must occur. Thus the advent of perturbation kinetics, developed in the 1950s, occurred and led to Nobel awards for the leaders of two different laboratories. The essential insight of perturbation kinetics was that a solution could be prepared at some temperature and pressure, and then a perturbation applied in a very short time that would change conditions in the sample uniformly, initiating a reaction that would begin simultaneously throughout the entire volume of the mixture. The two implementations of perturbation kinetics were different in detail and spawned separate traditions.

External Perturbation. In Germany, Manfred Eigen and his collaborators demonstrated how to carry out kinetic measurements by perturbing an external thermodynamic parameter, such as temperature, *T* (2). In a *T*-jump experiment, the temperature of the sample, usually a liquid solution, is increased suddenly by about 5°C and maintained at that new temperature, while the kinetics of the response of the mixture is measured. Usually the time required to make the kinetic measurement is short enough that maintaining the new temperature is not an issue; but some investigations have incorporated active stabilization at the new, higher temperature. Changing *T* often has a significant effect on the equilibrium constant, *K*_e. A factor of two or three change is possible over 10°C. The original set of concentrations existing before the *T*-jump is not the set of equilibrium concentrations appropriate to the new temperature. The kinetic measurement involves observing the relaxation of the concentrations toward the values appropriate to the new temperature. This procedure requires that equilibrium not lie too far in either direction and that there be sufficient dependence of *K*_e on *T*, but is otherwise quite general.

For the simple, prototypical reaction



if the equilibrium concentrations at the new temperature are *A*_∞, *B*_∞, and *P*_∞, assuming that raising *T* drives the equilibrium toward the left, encouraging dissociation, then prior to the *T*-jump and immediately after, before relaxation, the initial concentrations are *A*_i = *A*_∞ − Δ*A*(0), *B*_i = *B*_∞ − Δ*B*(0), and *P*_i = *P*_∞ + Δ*P*(0). The kinetic measurement involves the time evolution of Δ*A*(*t*), Δ*B*(*t*), and Δ*P*(*t*). The governing differential equation for [*P*] might be assumed to be

$$\frac{d[P]}{dt} = k_f[A][B] - k_r[P]$$

13

or, equivalently,

$$\frac{d[P_\infty + \Delta P]}{dt} = k_f[A_\infty - \Delta A][B_\infty - \Delta B] - k_r[P_\infty + \Delta P]$$

14

Three simplifications may be invoked: *d*[*P*_∞]/*dt* = 0; *k_r*[*A*_∞][*B*_∞] − *k_r*[*P*_∞] = 0, both from the definition of equilibrium; and *k_f*[Δ*A*][Δ*B*] may be assumed to be negligible because it is the product of small factors. Consequently, a simple linearized equation results:

$$\frac{d[\Delta P(t)]}{dt} = k_f([A_\infty][\Delta B] + [B_\infty][\Delta A]) - k_r[\Delta P]$$

15

From the stoichiometry of equation 12, the initial deviations have equal magnitudes, [Δ*A*] = [Δ*B*] = [Δ*P*] and

$$\frac{d[\Delta P(t)]}{dt} = \{k_f([A_\infty] + [B_\infty]) - k_r\} \Delta P(t)$$

16

The solution of equation 16 is a decreasing, simple exponential where *k*_{obs} = *k_f*([*A*_∞] + [*B*_∞]) + *k_r*. The perturbation approach generates small deviations in concentrations that permit use of the linearized differential equation and is another instance of pseudo-first-order behavior. Measurements over a range of [*A*_∞] + [*B*_∞] allow the kineticist to plot *k*_{obs} against that quantity and determine *k_f* from the slope and *k_r* from the intercept.

A number of approaches have been demonstrated for producing a *T*-jump. A compact and inexpensive strategy uses Joule heating by an electrical discharge through a conducting salt solution (Fig. 2). The energy stored in a capacitor is *E* = *CV*²/2. A 0.1 μF capacitor charged to 20 kV stores 40 J (10 kcal), which is enough to raise the temperature of 1 mL of water solution about 10°C. If the resistance of the solution, *R* = 20 Ω, and the resistance of the rest of the circuit is negligible, then almost all of the energy should be deposited in the salt solution in a few microseconds after the switch is closed. The capacitor is recharged over several seconds through the high value resistor shown. Because the characteristic time for discharge of a capacitor is τ = *RC*, assuming inductance is negligible, *C* and *R* must both be small in order to minimize the discharge time. Small *C* implies that voltage *V* must be large to store enough energy, which requires careful engineering to avoid inadvertent dielectric breakdown and ensure operator safety. Small *R* implies large salt concentrations. This, in turn, limits the choice of solvents, and in practice, water is used almost exclusively, making the method a natural for some important reactions, such as proton transfer (acid–base) reactions, which are often very fast, and many biochemical reactions. High voltage discharge precludes conductivity measurements, so uv–vis absorption spectroscopy is the rule for detection, and that is compatible with many reactions in those categories. Very short discharge times encounter another problem, however. When the temperature is raised, the solution normally expands. Density

variations lead to acoustic waves and even cavitation, which have very deleterious effects on light transmission. The problem can be minimized by working near a temperature at which the coefficient of expansion of the solution is zero, which occurs for aqueous solutions near 4°C. On the other hand, this temperature may not be the most interesting and precludes detailed investigation of temperature dependences. Pressurizing the solution also helps. Perturbation methods produce small concentration changes, which means that data are noisy. In the T -jump experiment the reaction is usually reversible, and measurements may be repeated on the same sample after it has cooled down again to the initial temperature. Extensive signal averaging is not convenient, because of the slow cycle time needed for cooling, but a dozen measurements can be combined in a reasonable time.

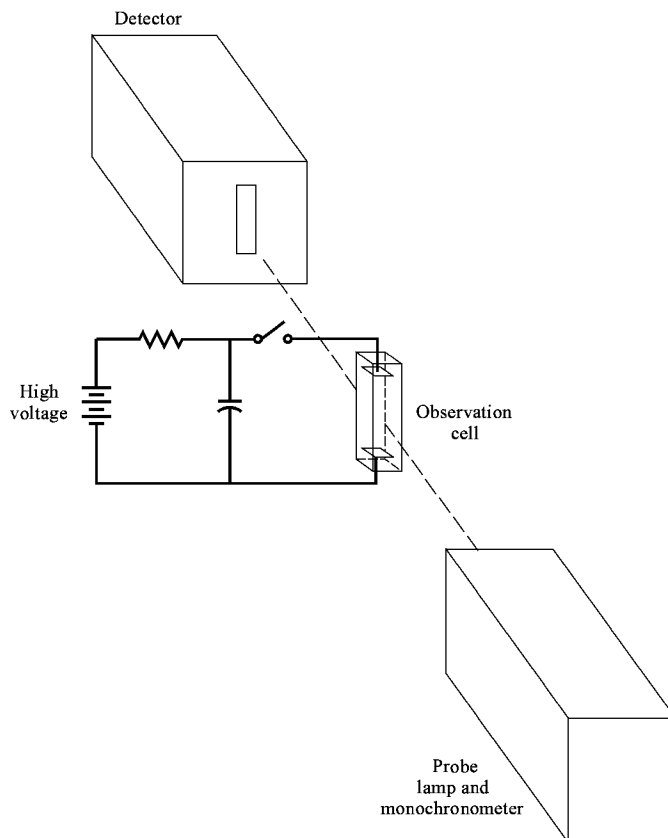


Fig. 2. Schematic of apparatus for temperature-jump (T -jump) measurements.

Liquid solutions can be heated by other means. Microwave irradiation was used very early for T -jump (see [MICROWAVE TECHNOLOGY](#)). The method giving the fastest heating times has been irradiation using light from a Q-switched, giant pulse laser (3), which can deposit energy in about 10 ns (see [LASERS](#)). The laser energy may be absorbed directly by the solvent. Water absorbs certain wavelengths in the near ir, and reactants are unlikely to absorb there. Thus undesired photochemical side reactions are avoided. Alternatively, an inert species that absorbs the laser radiation but does not interfere with the reaction or the monitoring of concentration changes may be added. Consequently, there is no restriction to aqueous solutions. The primary shortcoming is the limited energy available in a laser pulse. Depositing even 1 J uniformly throughout the solution requires a very expensive laser system, and 1 J produces only a T -jump of 0.25°C mL of water. Small volumes are needed.

Other perturbations have been demonstrated. The pressure, p , jump, similar to the T -jump in principle, is attractive for organic reactions where Joule heating may be impractical both because of the solvent being used and because concentrations might have to be measured by conductivity. Large (10^4 – 10^5 kPa) pressures are needed to perturb equilibrium constants. One approach involves pressurizing a liquid solution until a membrane ruptures and drops the pressure to ambient. Electric field perturbations affect some reactions and have also been used (2), but infrequently.

The perturbation methods described induce change as a step function, where the transition is much faster than the kinetics to be measured. There is no necessity for such a restriction. If an oscillatory temperature or pressure variation can be applied to a sample, reactant concentrations attempt to follow the oscillatory perturbation. If the perturbation oscillates slowly, the concentrations have a maximum amplitude and remain in phase with the perturbation. If the perturbation is introduced at a higher frequency, faster than the chemical kinetics can follow, there is a phase lag and a reduced amplitude for the oscillatory concentration changes occurs. Methods based on this approach facilitate extensive signal averaging. Rapid temperature changes of an oscillatory nature are difficult to impose, but there has been some demonstration of kinetic measurements using the lag of concentrations behind a monotonic increasing temperature change. This amounts to the same principle, restricted to less than one cycle, but is not particularly useful.

A periodic pressure change is attractive, compared with bursting membranes. A version closely allied with the classic p -jump has been demonstrated (4) and has been gaining in popularity. An oscillatory pressure wave is also a sound wave, and acoustic perturbations are used in conjunction with a different detection strategy. Instead of monitoring concentrations, the attenuation of ultrasound waves as a function of frequency are measured (2). Absorption of sound is maximized at frequencies corresponding to chemical relaxation rates. Acoustic methods remain a specialized procedure, however, because very different apparatus is needed for different time regimes. An extreme case of pressure perturbation is shock wave propagation, which has a separate history in kinetic studies of very fast reactions. Electric field perturbations are also well suited to sinusoidal modulation.

Flash Photolysis and Pulse Radiolysis. In England, Ronald Norrish and George Porter shared the Nobel prize for the development and application of flash photolysis (5). Unlike the perturbation methods, flash photolysis leaves pressure and temperature unchanged, using a flash of light to perturb reactant concentrations. Both irreversible and reversible reactions may be studied. In the former, the light flash is absorbed by an unreactive precursor to generate active reagents, the subsequent kinetic behavior of which is monitored. The precursor must be replaced before the measurement can be repeated. In the latter case, the perturbation produces an active species that starts the chemical change of interest, but ultimately the system relaxes back to the original equilibrium, from which it may again be perturbed. A short flash of visible or uv light is a convenient excitation. Shorter wavelength radiation in the x-ray region would be similar in principle, but is used only when there is an interest in the effect of x-rays or gamma rays and not as a general means of starting chemical reactions. Closely related is pulse radiolysis. This method is generally understood to imply that a particle beam, usually electrons, is incident on the sample and starts the reaction, either by supplying electrons or protons as a reagent, or by affecting a precursor species. In all of these cases, a single quantum of excitation, be it photon or particle, carries energy greatly in excess of characteristic thermal energies, kT , and sufficient to break a chemical bond or otherwise have a significant influence on the molecule excited.

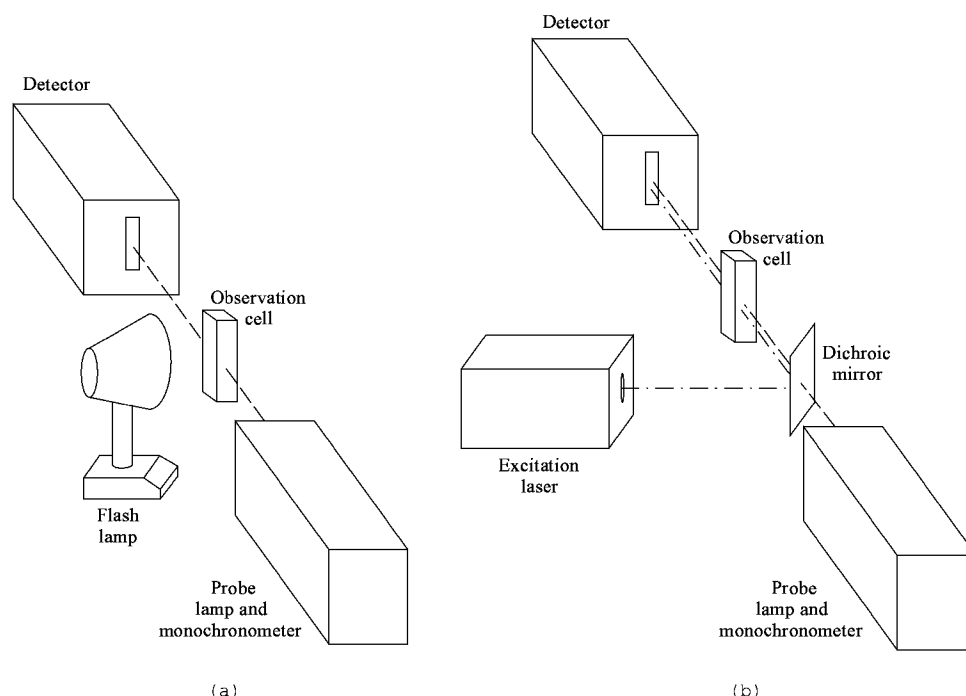


Fig. 3. Schematic of apparatuses for flash photolysis. (a) A simple instrument, and (b) a more sophisticated one utilizing longitudinal excitation.

Flash photolysis is prominent in studies of photochemistry (see [PHOTOCHEMICAL TECHNOLOGY](#)). Important examples of photochemical processes, both desirable and undesirable, occur in natural phenomena, such as photosynthesis or the induction of skin cancer. Over the past century, photochemical processes have assumed increasing importance in applied technology, such as photography (qv) and optical data storage (see [INFORMATION STORAGE MATERIALS, OPTICAL](#)). There are unique experimental advantages of flash photolysis. A primary force driving early developments in flash photolysis was interest in studying organic reactions to detect the very reactive free-radical species suspected to exist as reaction intermediates in thermally induced reactions.

Figure 3a shows the simplest arrangement for flash photolysis studies. The optical flash is introduced perpendicular to a spectrophotometric probe beam. It is usually necessary to restrict the excitation to a limited wavelength region and to isolate the detector from scattered exciting light. For measurements down to milliseconds, the flashlamp excitation can be as simple as a common photographic flashgun. The probe is usually optical absorption, as shown, but could be any on-line analytic technique. For times down to about one second, any modern laboratory spectrophotometer could be adapted to monitor concentrations following flash excitation. For shorter times, simple analogue-to-digital conversion boards plugged into personal computers are used. Historically, an oscilloscope was used, but computers are both more convenient and less expensive. At millisecond or longer times, the detection part of flash photolysis is similar to methods used for external perturbations.

The unique virtue of flash photolysis is that it can be extended another 11 orders of magnitude toward shorter times. Down to times of a few nanoseconds, the most common procedures employ essentially the same principles used for millisecond experiments. The same excitation energy is delivered in a shorter flash, and faster electronics are used to monitor changes in a continuous probe of concentrations. The probe is modified as necessary to permit faster measurements, using a brighter lamp to maintain an adequate signal-to-noise ratio while measuring faster transmittance changes. In the pre-laser era, flash photolysis technique developed in the direction of generating very energetic excitation flashes capable of making substantial concentration changes throughout a large volume so that kinetic changes could be monitored in a single flash. Signal averaging was rarely employed, except when using certain luminescence methods. It was difficult to make bright excitation flashes shorter than a microsecond, except for luminescence. Once pulsed lasers became available, pulse durations of 10 ns were easily attained. Several technologies, such as Q-switching, pulsed electrical excitation of gas discharge lasers, including the important uv-emitting excimer lasers, cavity-dumping, and even pulsed excitation of semiconductor lasers, conveniently generate pulses having durations near 10 ns. The first two methods can produce pulses with hundreds of millijoules of energy at rates of 1 to 1000 Hz; the last two generate smaller energy pulses at repetition rates ranging from 1 to 1000 kHz.

In addition to producing short pulses of spectrally well-defined light, lasers have the advantage that their optical energy can be delivered to the sample efficiently and focused to illuminate a small volume. In order to produce large transmittance changes, the energy should be delivered so as to create a long optical pathlength for the probe with a small cross-sectional area. The geometry of Figure 3b shows longitudinal excitation, which accomplishes this goal. In some cases, truly microscopic methods can be used. An energy of 1 μJ deposited in a volume less than 0.01 mm^3 yields an energy density greater than 0.1 J cm^{-3} , which can be sufficient to produce a millimolar concentration of chemical transients. Such microjoule energies are compatible with high repetition rates and extensive signal averaging to attain impressive sensitivity. Repeated excitation of a small volume, however, often leads to sample degradation, which can be avoided by using flow systems for gas or liquid solutions, and spinning or translating samples for solids.

Nanosecond laser pulses are also compatible with the fastest conventional electronic methods. Furthermore, such times are short enough for the fastest bimolecular reaction processes. Because there is no upper limit to the times that may be investigated, laser flash photolysis spans the entire time range of ordinary chemical kinetics using a single instrument. The probe apparatus, however, should be optimized for different time regimes. Times shorter than about a microsecond require very bright probe lamps in order to resolve changing concentrations. Pulsed flashlamps are used because these are bright while the measurement is made, but require less electrical input power, reduce cooling requirements, and minimize needless exposure of the (presumably photosensitive) sample to light. Lasers are an attractive alternative for the probe, but are still not easily tunable over wide wavelength ranges. Bright probe lamps also require that extra attention be given to detectors (see [PHOTODETECTORS](#)). Photomultipliers can introduce subtle nonlinear distortions in data long before they fail in any obvious manner. Specially designed high current photomultipliers and associated electronic circuits are used, but large-area, high current avalanche photodiodes are expected to be the wave of the future. For nanosecond times, electronic digitization cannot be accomplished by a simple plug-in card in a microcomputer. Instead, an intermediate instrument acts as a buffer to capture the transient signal and store it temporarily before transfer to a general-purpose microcomputer.

Analogue oscilloscopes having photographic recording, once common, are tedious and no longer cost-effective in any time range. Also vanishing are sampling methods which require that the measurement cycle be repeated identically many times. Although appropriate for repetitive electronic signals, sampling is not a good strategy for chemical kinetics. Even if the excitation can be repeated millions of times, the chemical system cannot, in general, tolerate so much photolysis. Recording digitizers sample at intervals as short as 0.2 ns, while measuring thousands of time samples for each excitation flash.

Whereas traditionally most kinetic measurements techniques monitored changes at a single wavelength, electronic methods for measuring hundreds of wavelengths at a particular time delay are evolving. Diode arrays and charge-coupled devices (CCDs) are revolutionizing the design of spectrographs, combining the multiplex advantages of photographic procedures with the convenience of electronic detection. Rapid scan spectrophotometers can measure complete spectra every millisecond. The most profound advance in fast kinetic studies, however, is progress in using vibrational analyses for the probe,

both time-resolved ir absorption and time-resolved Raman scattering.

Very Fast Kinetics. One nanosecond is by no means the limit for kinetic measurements. The state-of-the-art for direct, time-domain measurements lies close to 10^{-14} s (10 fs). At times shorter than a few nanoseconds, however, there is usually not a well-stirred solution. Even using a uniform distribution of precursor molecules and a homogeneous excitation pulse, the reagents created by photolysis retain the memory that they originated from precursors. Each might recombine with its partner fragment to reform the precursor in a process called geminate recombination (from *geminate*, meaning twin or born together). Another concern is orientational randomness. Photolysis creates a distribution of reactants that is partially aligned. Large molecules, such as proteins (qv), take nanoseconds to become rotationally randomized.

Time-domain kinetic measurements in the picosecond and femtosecond regimes rely almost exclusively on mode-locked lasers. Whereas other pulsed lasers generate nanosecond pulses by, in some sense, turning the lasing action on and off, mode-locked lasers depend on a more subtle phenomenon. These latter use constructive and destructive interference of a superposition of light waves to produce short spikes of light, which repeat at intervals to produce a train of pulses that are usually spaced about 10 ns apart. Each pulse has a duration ranging from 30 ps down to 10 fs. The pulse train can be used directly, or some of the picopulses can be extracted and amplified to larger energies. Different amplifier designs and different strategies for signal averaging are appropriate for 10 Hz, 1 kHz, and 100 kHz systems. The average power was about the same, 10 mW, throughout the 1980s regardless of the repetition rate. In the 1990s, 1 mJ pulses have become available at 1 kHz repetition rates, giving an average power of 1 W. This advance became possible with the development of a new generation of solid-state lasers, particularly the titanium–sapphire laser.

Picosecond kinetic studies rely on different measurement strategies from those used at longer times. A stroboscopic strategy is usually adopted. Short laser flashes are used not only to initiate the reaction, but also to probe concentrations at subsequent times. This is a sampling method, measuring one time delay per excitation pulse, and is used out of necessity, despite the general principle to the contrary. In compensation, it is common to measure an entire spectrum of wavelengths with each excitation cycle. The time delay between excitation and probe is determined by changing the position of mirrors to introduce a variable delay into the optical path. Because the reciprocal of the speed of light is about 3.3 ps/mm, micrometer adjustments readily give subpicosecond time delays. A meter of delay extends measurements into the nanosecond regime. Almost always the probe pulse is generated by extracting a fraction of the excitation pulse and, if necessary, using nonlinear optical processes to shift the wavelength (see NONLINEAR OPTICAL MATERIALS). The optical detection and electronic digitization simply respond to the total light in the probe pulse and have nothing directly to do with the time resolution.

Other Measurement Strategies

Combined Methods. Combinations of the procedures described herein are useful, especially for the study of unstable reagents. Double stopped-flow is increasingly common. Two reagents are mixed in the usual way and allowed to incubate for, eg, one second, to produce the reactive but unstable reagent of interest. Then the mixture is combined in the measurement cell with the contents of a third syringe to initiate the reaction of interest, which is studied in the usual way. In a similar vein, components may be combined using stopped-flow and the resulting mixture investigated by a *T*-jump timed to occur soon after mixing is complete. Flash photolysis could be done in the same manner, but because flash photolysis lends itself to signal averaging at high repetition rates, a continuous flow mixer with a flash photolysis cell located downstream at the appropriate time delay works very well.

Indirect and Novel Methods. Each incremental advance in time resolution is described in the kinetics literature as a first direct measurement of Often comparable data were available earlier by other methods. What counts as direct or indirect is often a matter of perspective. A direct measurement can be considered a record of changing concentrations as a function of time. Some researchers are more comfortable when the same information is Fourier transformed into the frequency domain. Since the latter 1800s it has been possible to measure the absorption of electromagnetic radiation as a function of frequency (or wavelength) and interpret lineshapes to yield kinetic information on the picosecond time scale. In uv–vis and ir spectra of gases, linewidths increase at high pressure owing to collision broadening. Dissociation or ionization may also determine a linewidth. More subtle effects occur in liquids. More recently, chemical reactivity on slower time scales has been measured by lineshape analysis in Mossbauer spectroscopy and magnetic resonance. Although nmr can be used simply to monitor concentrations at different points in time, usually over a span of some minutes, linewidth studies are more prominent in nmr, where measurement of exchange rates, the inverse of the mean lifetime of some species that undergoes association or dissociation, is more usual. Such linewidths are now actually measured in the time domain, confirming the time–frequency duality. Better than any other common procedure, nmr can distinguish one atom within a complex molecule and answer detailed questions about its behavior. Consequently, nmr has assumed a steadily increasing role in kinetic studies since its inception. This role is expected to expand as new techniques facilitate the study of larger species such as proteins. A specialized vocabulary is used in nmr kinetic studies (see MAGNETIC SPIN RESONANCE).

To reinforce the connection with time–domain kinetics, linewidth studies can be understood as monitoring the spontaneous fluctuations in concentrations that occur when a chemical system is at equilibrium:

$$AB \rightleftharpoons C \rightleftharpoons AC \rightleftharpoons B$$

(17)

The concentration $[AB]$ constantly experiences tiny fluctuations, the duration of which can determine linewidths. It is also possible to adopt a traditional kinetic viewpoint and measure the time course of such spontaneous fluctuations directly by monitoring the time-varying concentration in an extremely small sample (6). Spontaneous fluctuations obey exactly the same kinetics of return to equilibrium that describe relaxation of a macroscopic perturbation. Normally, fluctuations are so small they are ignored. The relative amplitude of a fluctuation is inversely proportional to the square root of the number of *AB* entities being observed. Consequently, fluctuations are important when concentrations are small or, more usefully, when volumes are tiny.

Another trend in analytic measurement is converging with kinetic studies and should force consideration of fluctuations. In the realm of microscopy (qv), spontaneous fluctuations have long been recognized in the behavior of objects that are larger than single molecules but still small, such as diffusion of Brownian particles or electrical conduction in nerves. Although electron and ion microscopy have molecular resolution, these samples are dead and unchanging. Newer techniques, eg, scanning tunneling microscopy, atomic force microscopy, and near-field optical microscopy, may offer molecular-scale resolution for functioning systems. Routine detection of single small molecules is close, and kinetic studies should follow. Of course, it will then be necessary to average over many single molecules to build up statistical information.

Even without resolving single molecules, optical microscopy is undergoing a revolution. The advent of sophisticated electronic detectors that facilitate quantitative treatment of changes observed in very small volumes brings kinetics to microscopy. In some cases, perturbation methods can be developed to facilitate kinetic studies within selected, small volumes. One powerful strategy in biology involves designing a chemical precursor that can be infused into a specific cellular region, where it remains inert until a sudden flash of light modifies it to an active form. The kinetics of subsequent behavior can then be followed either by traditional methods, usually fluorescence, or by recording and analyzing a motion picture of structural changes. It is not easy to devise useful precursors, however, and fluctuations increase as systems of interest become ever smaller.

Just as chemical structure determinations relying on physical methods have largely replaced structure proofs based on chemical reasoning, rapid instrumental measurements have largely, but by no means completely, supplanted an earlier tradition that relied on a more chemical strategy of measuring concentrations of reactants or intermediates by intercepting these with a scavenger that reacted quickly to form a stable product that could be quantified later. Such trapping methods can be construed as being indirect in the sense that they rely on one chemical reaction being faster or slower than another.

Experimental Variation of Chemical Rates with Temperature and Pressure

The experimentally measured dependence of the rates of chemical reactions on thermodynamic conditions is accounted for by assigning temperature and pressure dependence to rate constants. The temperature variation is well described by the Arrhenius equation,

$$k_c = A(T) \exp(-E_a/RT)$$

(18)

The prefactor $A(T)$, also called a frequency factor, has units of inverse seconds. It may have a weak dependence on temperature. Some theoretical models predict a variation with $T^{1/2}$, but such variation is frequently ignored and A is taken as constant over limited temperature ranges. The prefactor A is often

found to lie in the range of 10^9 to 10^{13} . The parameter E_a is termed the activation energy, and commonly ranges from zero up to a few hundred kJ mol⁻¹, comparable to the energy of moderately strong chemical bonds. The crude but oft-quoted estimate that a 10°C increase in temperature doubles a reaction rate is valid when E_a is 90 kJ mol⁻¹ (21.5 kcal mol⁻¹). When A is assumed constant, the activation energy is determined from a plot of

$$\frac{d(\ln(k_c))}{dT} = -\frac{E_a}{RT} \tag{19}$$

Increased temperatures should always increase the rates of reactions. Pathological exceptions would probably involve changes in the molecular species involved.

The dependence of k_c on pressure, rarely measured until the 1980s, is considered an important part of a comprehensive kinetic study. Gas-phase reactions, of course, depend strongly on pressure through the dependence on concentrations, which change with the partial pressure of reagents. The variation of k_c with pressure is a different and much smaller effect. A factor of two change in rates for common liquid solutions might require an increase to as much as 100 MPa (1000 atm). General considerations and precautions for working at high pressures apply to kinetic studies (see HIGH PRESSURE TECHNOLOGY). Almost all of the important kinetic methods, including stopped-flow mixing, T -jump, flash photolysis, and nmr methods, have been adapted for use at high pressures (7). The dependence on pressure is characterized by an activation volume, $V_a^\ddagger$, where

$$\frac{d(\ln(k_c))}{dp} = -\frac{V_a^\ddagger}{RT} \tag{20}$$

Activation volumes may be either positive or negative, corresponding to the fact that reactions may become either slower or faster at high pressures.

Microscopic Models in Kinetics

Mechanisms. Mechanism is a technical term, referring to a detailed, microscopic description of a chemical transformation. Although it falls far short of a complete dynamical description of a reaction at the atomic level, a mechanism has been the most information available. In particular, a mechanism for a reaction is sufficient to predict the macroscopic rate law of the reaction. This deductive process is valid only in one direction, ie, an unlimited number of mechanisms are consistent with any measured rate law. A successful kinetic study, therefore, postulates a mechanism, derives the rate law, and demonstrates that the rate law is sufficient to explain experimental data over some range of conditions. New data may be discovered later that prove inconsistent with the assumed rate law and require that a new mechanism be postulated. Mechanisms state, in particular, what molecules actually react in an elementary step and what products these produce. An overall chemical equation may involve a variety of intermediates, and the mechanism specifies those intermediates. For the overall equation



this two-step mechanism might be postulated



where I is an intermediate. In this case, the single arrow is meant to specify that equilibrium lies far to the right. A source of confusion is that mechanisms and stoichiometric relations appear the same when written as equations, despite the very different connotations. A complete mechanism implies a set of differential equations. Assuming well-stirred conditions, the above pair of equations predicts

$$\frac{d[I]}{dt} = k_f[A][B] - k_r[I] \tag{23a}$$

$$\frac{d[P]}{dt} = k_2[I][C] \tag{23b}$$

Much of the language used for empirical rate laws can also be applied to the differential equations associated with each step of a mechanism. Equation 23b is first order in each of I and C and second order overall. Equation 23a implies that one must consider both the forward reaction and the reverse reaction. The forward reaction is second order overall; the reverse reaction is first order in $[I]$. Additional language is used for mechanisms that should never be applied to empirical rate laws. The second equation is said to describe a bimolecular mechanism. A bimolecular mechanism implies a second-order differential equation; however, a second-order empirical rate law does not guarantee a bimolecular mechanism. A mechanism may be bimolecular in one component, for example $2A \rightarrow I$.

The solution of the simultaneous differential equations implied by the mechanism can be expressed to give the time-varying concentrations of reactants, products, and intermediates in terms of increasing and decreasing exponential functions (8). Expressions for each component become complicated very rapidly and thus approximations are built in at the level of the differential equations so that these may be treated at various limiting cases. In equations 2222 and 2323, the first reaction may reach equilibrium for $[I]$ much more rapidly than I is converted to P . This is described as a case of pre-equilibrium. At equilibrium, $k_f[A][B] = k_r[I]$. Hence,

$$[I] = K[A][B] \tag{24a}$$

where $K = k_f / k_r$, and

$$\frac{d[P]}{dt} = k_2 K[A][B][C] \tag{24b}$$

Experimental tests of this mechanism can determine the reaction order with respect to each component and verify the molecularities assumed, but are unable to separate even the factors k_2K , let alone measure k_f and k_r , as long as the assumption of pre-equilibrium remains valid. Better time resolution in the experiment captures the approach of $[I]$ toward equilibrium and, consequently, violates that assumption.

A second common approximation is the steady-state condition. That arises in the example if k_2 is fast compared with k_r in which case $[I]$ remains very small at all times. If $[I]$ is small then $d[I]/dt$ is likely to be approximately zero at all times, and this condition is commonly invoked as a mnemonic in deriving the differential rate equations. The necessary condition is actually somewhat weaker (9). For equations 22a and b, the steady-state approximation leads, despite its different origin, to the same simplification in the differential equations as the pre-equilibrium condition, namely, equations 24a and b.

Rate constants and their activation parameters are given microscopic interpretations. At the simplest level, a large E_a is interpreted by postulating a large potential barrier between reactants and products that must be overcome by kinetic energy available in the reactants, which increases at higher temperatures. Similarly, a positive $V_a^\ddagger$ in a liquid-phase reaction, which corresponds to a reaction that slows at high pressure, is interpreted by postulating that the reactants pass through some configuration that has a larger volume on the way to forming products. Increasing the pressure makes it more

difficult to effect that expansion against the surrounding medium. At more advanced levels of analysis, detailed atomic models are developed which define microscopic concepts that attempt to predict activation parameters. These models may be referred to as activated complex theory, transition-state theory, absolute rate theory, or any of a variety of terms. The theoretical models may be rather ad hoc, or these may be grounded more or less rigorously in quantum statistical mechanics. In any case, the activation energy barriers and volumes are assigned to transition states, not to intermediates. The latter are real chemical species having a finite, but possibly very short, lifetime, whereas the former have only the most fleeting existence at some set of atomic positions. With the development of femtosecond laser measurements, the challenge to refine exactly the meaning of such distinctions has arisen. There can be heated debates about interpreting quantities such as activation parameters in atomic terms. It is therefore important to distinguish numbers that simply parametrize experimental data, and should be subject to no debate beyond accuracy and precision, from model-dependent interpretations of those numbers.

Reaction Dynamics. Mechanism is to chemical change as structure is to chemical identity. Throughout the twentieth century the goal of structural chemistry has been to specify accurate spatial coordinates for all the atoms in a molecule or crystal. In contrast, the mechanisms that have been the goal of kinetic studies are much less detailed. As of this writing (ca 1994), there is some progress being made toward a detailed atomic level description. This is such a profound advance that practitioners do not describe themselves as doing chemical kinetics, but prefer a name such as reaction dynamics (10). In a crude sense, the goal is to specify all atomic positions as a function of time, but the treatment must be grounded firmly in quantum theory, which limits such a description. Early progress toward these goals was recognized with the award of a 1986 Nobel Prize to Dudley Hirschbach and Yuan Lee of the United States and John Polanyi of Canada. For the simplest cases, such as the reaction of a small diatomic, eg, dihydrogen, and a single small atom, eg, deuterium, there is a detailed quantum mechanical treatment that specifies initial and final states and everything in between that nature allows to be specified. For cases that are slightly more complicated, the full quantum treatment is not yet available, but experiment and theory both strive to define exact quantum states for reactants and products and offer as detailed a description as possible for the transformation. Once a description for all relevant quantum states of reactants has been measured or calculated, a suitable average should be performed to regenerate the rate constants that are the goal of classical kinetics. In practice, that is not likely and may never be useful for most applied purposes. The principal practical consequence of the newer insights are expected to be a clarification of how and when it is appropriate to invoke a rate constant.

Detailed reaction dynamics not only require that reagents be simple but also that these remain isolated from random external perturbations. Theory can accommodate that condition easily. Experiments have used one of three strategies. (1) Molecules in a gas at low pressure can be taken to be isolated for the short time between collisions. Unimolecular reactions such as photodissociation or isomerization induced by photon absorption can sometimes be studied between collisions. (2) Molecular beams can be produced so that motion is not random. Molecules have a nonzero velocity in one direction and almost zero velocity in perpendicular directions. Not only does this reduce collisions, it also allows bimolecular interactions to be studied in intersecting beams and increases the detail with which unimolecular processes that can be studied, because beams facilitate dozens of refined measurement techniques. (3) Means have been found to trap molecules, isolate them, and keep them motionless at a predetermined position in space (11). Thus far, effort has been directed toward just manipulating the molecules, but the future is bright for exploiting the isolated molecules for kinetic and dynamic studies.

In comparison with isolated molecules, reactions in liquids, or on surfaces, or even in dense gases are difficult to treat, and methods are still at a very early stage. Nevertheless, there is a spirit of optimism based on the threefold combination of fast computers able to handle detailed models, such as molecular dynamic simulations; ultrafast measurement techniques, which probe reactions on time scales as fast as the random surroundings can change; and insights of clever researchers in generating ever more realistic, but still tractable, descriptions of chemical change. Reaction dynamicists believe that they have a revolutionary vision that should supplant traditional kinetic treatments in the twenty-first century and supply chemists with details of reactions that consist essentially of a motion picture of chemical change with atomic detail (12).

BIBLIOGRAPHY

1.

P. Hanggi, P. Talkner, and M. Borkovec, *Rev. Mod. Phys.* **62**, 251–341 (1990).

2.

M. Eigen, *Disc. Faraday Soc.* **17**, 194–205 (1954).

3.

D. H. Turner and co-workers, *J. Am. Chem. Soc.* **94**, 1554–1560 (1972).

4.

R. M. Clegg and B. W. Maxfield, *Rev. Sci. Instrum.* **47**, 1383–1393 (1976).

5.

R. G. W. Norrish and G. Porter, *Nature* **164**, 658 (1949).

6.

D. Magde, E. Elson, and W. W. Webb, *Phys. Rev. Lett.* **29**, 705–708 (1972).

7.

R. van Eldik, ed., *Inorganic High Pressure Chemistry-Kinetics and Mechanisms*, Elsevier, Amsterdam, 1986.

8.

M. Eigen, in S. Claesson, ed., *Fast Reactions and Primary Processes in Chemical Kinetics*, 5th ed., Wiley-Interscience, New York, 1967, pp. 333–369.

9.

J. H. Espenson, *Chemical Kinetics and Reaction Mechanisms*, McGraw-Hill Book Co., Inc., New York, 1981, pp. 72–73.

10.

R. D. Levine and R. B. Bernstein, *Molecular Reaction Dynamics and Chemical Reactivity*, Oxford University Press, U.K., 1987.

11.

S. Chu, *Sci. Am.* **266**, N2:70–76 (1992).

12.

A. H. Zewail, *Sci. Am.* **263**, N6:76–82 (1990).

General References

Reference 8 includes articles by many of the founders of perturbations kinetics, written just as the field was maturing. K. Kustin, ed., *Fast Reactions, Methods in Enzymology*, Vol. 16, Academic Press, Inc., New York, 1969. Contains enough detail to allow one to build machines and make measurements. Predates lasers, fast electronics, and computers. H. Strehlow, *Rapid Reactions in Solution*, VCH, Weinheim, Germany, 1992. Recent review of perturbation kinetics and magnetic resonance methods. J. W. Moore and R. G. Pearson, *Kinetics and Mechanism*, 3rd ed., Wiley-Interscience, New York, 1981. The first edition by A. A. Frost and R. G. Pearson appeared in 1953 and the second edition by the latter authors appeared in 1961. Probably the best known graduate text in the United States that treats kinetics and mechanisms for general chemistry. J. I. Steinfeld, J. S. Francisco, and W. L. Hase, *Chemical Kinetics and Dynamics*, Prentice Hall, Englewood Cliffs, N.J., 1989. Oriented more toward gas-phase reactions and includes more advanced microscopic interpretations from the perspective called chemical physics.

Douglas Magde
University of California at San Diego

KRYPTON.

See HELIUM GROUP, GASES

KURCHATOVIIUM.

See ACTINIDES AND TRANSACTINIDES

LABELS.

See FLAME RETARDANTS; INDUSTRIAL HYGIENE

LABORATORY INFORMATION MANAGEMENT SYSTEMS

A laboratory information management system (LIMS) is a computer or computer network used to automate the acquisition and management of raw analytical data. In its simplest form, it tracks samples and test results through analytical laboratories and provides summaries of the status of these samples and tests. In its most advanced form, the system is interfaced to the laboratory's instrumentation and communication network to allow automation of data gathering, compilation, and reporting.

The first LIMS appeared in the 1970s, and their use became more widespread as a result of U.S. Federal Government regulations (1,2), particularly in the pharmaceutical industry. LIMS was seen as a method to meet the government agencies' requirement for laboratories to account for the results of analyses. In the mid-1980s, the U.S. Environmental Protection Agency (U.S. EPA) regulations for laboratories taking part in the Superfund program, to remediate designated sites, resulted in further expansion of the use of LIMS. In the early years, a LIMS was expensive and was customized for each installation using in-house personnel or a contract software vendor. More recently, the approximately 40 vendors selling LIMS have been able to design the systems with enough flexibility to meet the needs of most laboratories. Commercial systems can usually be delivered quickly and the laboratory benefits from a large user base, compatible accessories, and future enhancements.

The benefits of a LIMS depend largely on the needs of the laboratory and the type of system installed. In general, a LIMS improves the management of the laboratory by providing more accurate and timely information regarding work load and work-load distribution. The productivity is increased through the automation of many clerical and routine tasks associated with sample identification, tracking, and transcription of results. The quality of the data can be improved through automated data acquisition, reduction in transcription errors, and automatic enforcement of validation analysis and standardization procedures. Finally, the system can be used to meet regulatory agency compliance requirements.

LIMS Function

LIMS is a database management system designed to help laboratories, particularly analytical laboratories, manage the data generated in the laboratory. The LIMS database usually includes biographical information about the sample, such as its source, the customer, a project or charge number, and the final results of the analysis or analyses. An additional important function of a LIMS is to provide on-line information concerning a sample's status. This information can include the sample's location, the status of a particular analysis, and any completed results. LIMS is of most value when multiple or sequential analyses are required on one or a series of samples. A LIMS system can generate multiple requests and have samples split and sent to the appropriate laboratory. The system can then be used to compile results into a single report for the client. The various functions of a LIMS can be summarized as follows as analytical or managerial tasks (3) and Figure 1 outlines the typical flow of a sample coming into the laboratory (4). Suppliers of LIMS include Banyon Systems Inc., Beckman Instruments Inc., Challenger Group Inc., Chesapeake Software Inc., Cirrus Technology, Digital Equipment Corp., DSP Development Corp., Hewlett-Packard, IBM, Keithley Instruments, Laboratory Data Systems Inc., Laboratory MicroSystems Inc., LabWare Ltd., Northwest Analytical Inc., Novell Inc., PE Nelson Div., Radian Corp., Statistical Graphics Corp., 3Com Corp., Varian Associates Inc., and VG Instruments (5).

Analytical level tasks

- Sample number generation
- Bar-code label generation
- Sample log-in
- Verification of data format entered into the computer
- Worksheet generation
- Construction and checking of calibrated curves
- Direct data acquisition from chromatographs
- Data collection for analytical instruments
- Entry of instrumental readings
- Manual results entry
- Interpretation of calibrated curves and quality control samples
- Interpretation and acceptance of sample data
- Routine automatic calculations
- Plotting routines for visualization of analytical data

Managerial level tasks

- Acknowledgement of sample receipt
- Backlog investigation

- Sample and status tracking
- Database searches
- Numbers of samples assayed
- Tests utilized
- Numbers of samples analyzed per instrument
- Cost per assay
- Customer charges
- Results collation and presentation
- Report generation
- Scheduling and rescheduling of work
- Archiving and retrieval of data
- Workload status and the justification of equipment
- Regulatory agency compliance
- Audit trail for all database transactions
- Security: class or hierarchy
- Instrument records and calibration where appropriate

Automation

The development and widespread use of computers and microprocessors in control laboratory instruments has made it possible to fully automate a laboratory, including interfacing instruments directly to a LIMS. In the fully automated laboratory, a sample is logged into a LIMS, then transferred to a laboratory where it is prepared for analysis by a robot, which then transfers it to an autosampler or analyzer. Once analyzed, the data is transferred through a communications link to a device which could convert the raw data into information that a customer needs. For example, in a simple case, a nmr spectrum could be compared to spectra on file to yield an identification of an unknown. In more complex instances, a data set could be compared to standards and by using pattern recognition techniques the LIMS would be able to determine the source of a particular raw material. Once the data is reduced and interpreted, the LIMS becomes the repository of the information. A schematic for such a fully automated laboratory is shown in Figure 2 (6).

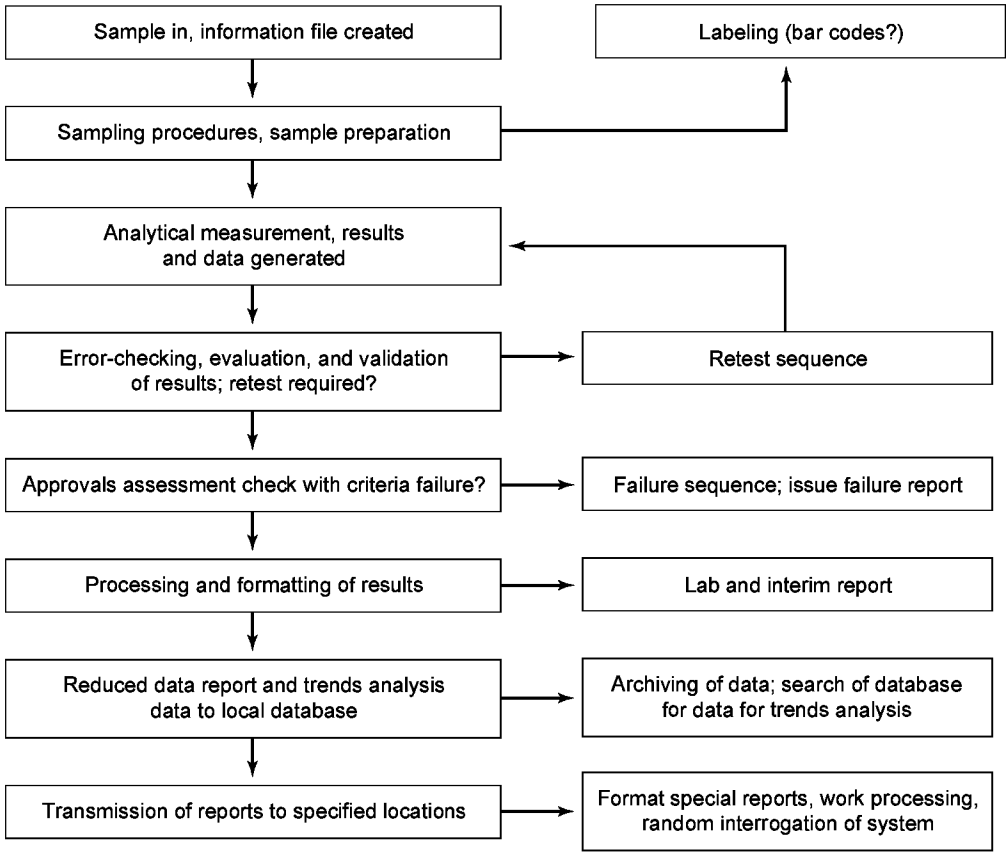


Fig. 1. LIMS procedure flow chart (sample handling and reporting) (4).

Courtesy of the Royal Society of Chemistry.

Quality Management

The business activity of the organization dictates quality requirements for the LIMS. Security and regulatory requirements for LIMS data define the level of effort expended to validate a LIMS and the data being stored. In addition, the quality of the hardware and software used to implement the LIMS both play a role in determining overall system quality.

The level of quality required in a LIMS is related to the particular role played by the system in the organization's business. In an R&D environment, for example, the primary reason for a LIMS might be to compare and document new analytical methods.

Manufacturing organizations often implement LIMS systems in their Quality Assurance (QA) and Quality Control (QC) laboratories. These labs ensure that the product is of high quality (QA), and examine intermediates to keep the production process on target (QC). Statistical quality control (SQC) uses techniques such as Shewhart control charts, Pareto charts, and time-to-failure (Weibull) analysis (7). Production labs incorporate such features as alarms to notify of SQC limit violations and trend violations. LIMS can be used in conjunction with computer-integrated manufacturing systems (CIMS) to pinpoint problems during production. Some LIMS systems have built-in statistical tools, and others enable the user to export data to other software packages for analysis.

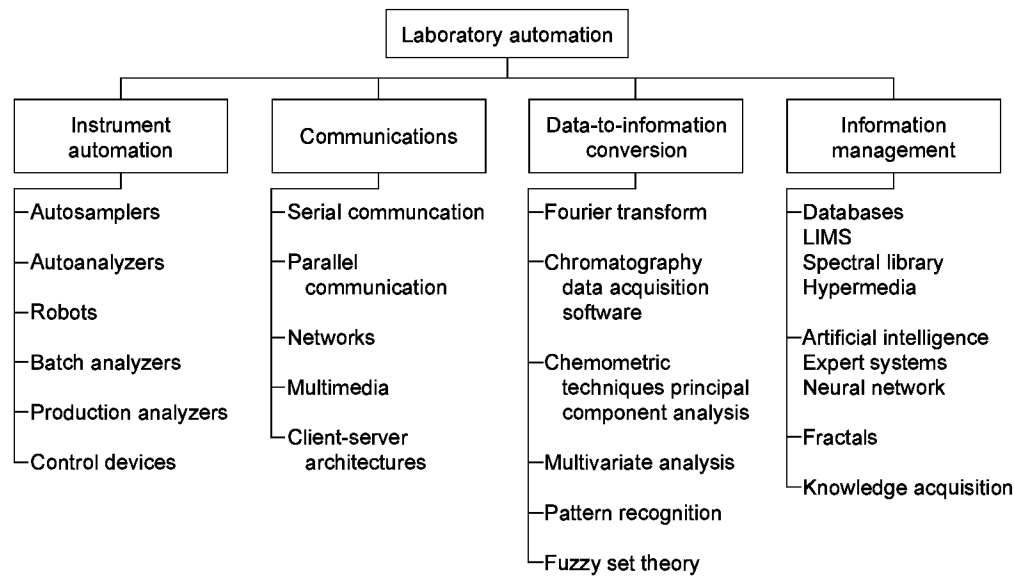


Fig. 2. Components of a fully automated laboratory (6).

Clinical laboratories in the pharmaceutical and medical fields are subject to regulatory and confidentiality requirements. The U.S. Food and Drug Administration issues guidelines for good laboratory practices (GLP) and good manufacturing practices (GMP). Organizations manufacturing regulated products for use in humans (such as drugs and medical devices) must validate the integrity of their LIMS systems. The EPA has drafted good automated laboratory practice (GALP) guidelines to be used in their contract laboratories. This emerging standard promises to become a worldwide framework for the validation of automated data integrity (8).

The International Standards Organization 9000 (ISO 9000) standard, developed by the European Economic Community (EEC), also impacts manufacturers' implementation of LIMS systems. ISO 9000 is a set of standards which are required for manufacturers selling products to the EEC. The 9000 standards are credited with playing an important part in the impetus to greater computerization of laboratory information management (9).

A LIMS system must be validated to ensure data security and integrity and to ensure it complies with applicable government regulations. The validation process includes first developing a strategy and setting specific objectives for validation. Modular testing of the system should be employed, and a detailed test protocol should be developed for each LIMS module (10). Documentation is critical throughout the validation period. The test strategies and procedures, as well as flow charts for the LIMS source code, must be documented. Training documentation for LIMS users and administrators should be included in a complete validation package.

The hardware and software used to implement LIMS systems must be validated. Computers and networks need to be examined for potential impact of component failure on LIMS data. Security concerns regarding control of access to LIMS information must be addressed. Software, operating systems, and database management systems used in the implementation of LIMS systems must be validated to protect against data corruption and loss. Mechanisms for fault-tolerant operation and LIMS data backup and restoration should be documented and tested. One approach to validation of LIMS hardware and software is to choose vendors whose products are precertified; however, the ultimate responsibility for validation remains with the user. Validating the LIMS system's operation involves a substantial amount of work, and an adequate validation infrastructure is a prerequisite for the construction of a dependable and flexible LIMS system.

Costs and Benefits

The first step in acquiring a LIMS is justifying a need for it. The investment must provide increased value to laboratory customers, while also providing benefits to LIMS users and their management. The organizational framework has to be adaptable to the operational changes that acquisition of a LIMS will cause. If computer technology is not widely used in the organization, it may be difficult to automate laboratory functions. LIMS systems are specialized information systems tailored to an existing framework of functions. If an information strategy has been defined for the organization, a LIMS can be justified in terms of how it fits in with that strategy. If the organization has computing standards, the LIMS chosen should conform to those standards whenever possible.

An excellent tutorial on evaluating the costs and benefits of LIMS has been presented (11). The tutorial explains how to perform a cost-benefit analysis using a pragmatic approach to the economics involved. To assist in the analysis, a detailed list of specific LIMS costs and benefits from the tutorial is included here (see Tables 1-3) (12-14). The size of the system, based on the number of samples and analyses per year, can be used to approximate the cost. Table 4 summarizes the respective costs for small, medium, and large systems (15). Although the tutorial is a good starting point for accurately assessing the costs of a proposed system, the task of measuring the value of the benefits can be much more subjective in nature.

Benefits can be classified as tangible and intangible. Tangible benefits are benefits that are easily assigned a monetary value. These include items such as reductions in the costs of calculating and reporting data, improved capacity through better access to management of data in making assignments, and being able to identify and report sample status more quickly. Intangible benefits might include better overall service to customers and a general perception of better laboratory management through the use of state-of-the-art management techniques.

Table 1. Initial LIMS Costs^a

Category	Cost factor
hardware	computer storage devices; device cabling and wiring; wiring, power conditioning; climate control; furniture; networking: interfaces, servers, cabling, wiring closets; user interaction: terminals, workstations; display devices: printers, plotters; sample ID: bar-code readers, printers, and label applicators; repair documentation, spares
software	operating system; system development tools, including communications and compilers; networking and connectivity software components; LIMS core and optional components; LIMS development tools; database systems and development tools; generic instrument interface software; special instrument interface software, eg, chromatography, process control; graphics and printing software; documentation: users, managers, development; furniture for storing documentation
installation and conversion	personnel to manage acquisition and installation; disruption due to installation; loss of replaced incompletely depreciated equipment; lost space for computers, terminals, printers; customization and configuration; development of computer usage procedures, LIMS usage procedures, and new

	laboratory procedures; determination off-line vs on-line data needs; development of archiving and backup procedures
training	computer training for users, managers, hardware engineers, support staff; LIMS training for users, managers, support personnel; instrument interface training for users, managers, hardware engineers, support staff; accessory software training for users, managers, support personnel; lost work during training; decrease in productivity during training, installation, and adaptation
configuration	time to identify database items; time to configure database to include these items; time to create necessary database forms; time to set up system security, desired audit trails, and enter valid users; time to design and create necessary reports, charts, and alarms; time to configure laboratory for tests, samples, instruments, projects, labs, etc; enter analysis procedures; set up accounting system, if used; develop interfaces to other corporate systems; hook up instruments, reconfigure automated instruments

^a Ref. 12.

Another significant benefit of a LIMS is the improvement of the overall quality of the laboratory. In the case of a laboratory, quality is defined as satisfying customer needs in the areas of accuracy, reliability, clarity, and timeliness of analytical information. LIMS can enhance quality in a number of ways, eg, in checking conformance to requirements, in organizing and prioritizing work to ensure timeliness, in measuring laboratory performance in areas of technical quality and efficiency so as to provide continuous improvement, and in helping the laboratory to communicate clearly, completely, and consistently (16).

Selection

Before deciding on a LIMS product, a complete set of specifications for required functions of the LIMS should be written. The best time to do this is before any vendors have been contacted. Inviting vendors to participate in this process results in a specification which can be automatically fulfilled by their product. A thorough understanding of current data management processes is necessary to draw up a specification (17), and vendors do not have it.

Table 2. Ongoing LIMS Costs^a

Category	Cost factor
supplies	printer paper, forms, toner, drums; plotter paper, pens; bar-code labels, ribbons; fuses, keys cables, spare parts; media for backups and archives; cleaning supplies
maintenance fees	computer and networking hardware; operating system; LIMS core and option software; database software; other software packages;
personnel	administration of maintenance contracts; installation of updates; performing and supervising in-house preventative and nonpreventative maintenance; performing and supervising backups and archiving; archival costs; internal support personnel time; training of new personnel due to normal promotions and turnover
miscellaneous	power; computer lease and rental costs; communication rental (telephone, satellite, etc); depreciation costs

^a Ref. 13.

When researching the requirements for a LIMS, answers to the following questions, courtesy of The Royal Society of Chemistry, should be sought (18). (1) Is the laboratory to be used for a single technique or multiple techniques, for a single product line or product type, or for all types of samples? Would a redesign of experimental procedures produce an increase in efficiency? (2) Are there management requirements, including secretarial assistance? (3) Are computing resources required, including a laboratory management data system, local data collection and storage of instrument results, data storage and filing requirements at laboratory level, use of personal workstations, introduction of local area networks (LANs), mainframe interfaces, and telecommunication facilities? (4) Are there reporting, archiving, and database requirements? (5) Are personnel records to be included in the computerized laboratory management scheme? (6) What overall response and speed of the system is necessary for the turnaround expected? (7) What is to be done about the education of users, a most important requirement that is easily overlooked in the planning stages?

The search for a vendor can begin once the specification is complete. The level to which a vendor's products conform to established requirements is the most important selection criterion for a LIMS. Financial stability of the vendor's organization is another consideration, because ongoing technical support from the vendor is vital to LIMS implementation and enhancements. The technology used in a vendor's products should be evaluated for robustness and longevity. For example, a vendor with software products written for nearly obsolete computing platforms would not be a strategic choice. Consulting services offered by a vendor can be valuable for implementing and maintaining a LIMS.

Table 3. LIMS Benefits^a

Category	Benefit factor ^b
data management	data may be stored more systemically; more data conveniently stored per event; track a greater number and more complex relationships; change data more quickly; data may be more secure; security can be more highly tailored and may require less managerial oversight; audit trails can be automated; transactions are more traceable; managerial information on the amount and type of data are improved; data can be recalled more quickly and flexibly; improved reliability of data storage and retrieval; data may be handled more flexibly; reporting is faster; unit printing costs and cost of transferring data may be reduced; new relationships among data are more easily examined; data are available when needed by requiring its entry
laboratory throughput	entry of sample information and sample identification may be faster; reduced transcription (offset by possible increased data entry time); faster QA QC or SPC; speed of analysis; automation of entire analysis; faster response to queries; faster and automated reporting; automatic validation of results; inspection only of discrepancies; assistance in remembering procedures, prompt for required data
quality of data	more accurate data because of automated acquisition and range checking; more accurate sample identification due to bar codes; reduced transcription errors; improved instrumentation reliability from QC and automated calibration procedures; more involved QC procedures; SPC on laboratory analyses; automatic enforcement of validation procedures, laboratory analysis procedures, and instrument standardization procedures; automatic prompt for missing data
laboratory management	improved accuracy of laboratory management information; increased quantity of laboratory management information; reduced labor to produce laboratory management information; more instruments, analysts, results handled per manager; reduction in lost opportunities due to inadequate knowledge and in unsolved

quality of operations	backlog problems; improved validation procedures
	SPC on manufactured products; SQC on laboratory operations; communicate with corporate CIM system; improved QA QC on products; reduced testing costs; correlate laboratory analyses and process measurements; faster solutions to production problems tested in laboratory; faster notification of backlog problems; improved electronic data interchange capabilities; automated communication with inventory, ordering, and materials planning systems
regulatory compliance	automated regulatory report generation; regulation-compliant audit trail; enforce regulation-compliant laboratory procedures and result validation; improved response to changing regulations; fast compliance with regulatory audits

^a Ref. 14.

^b QA — quality assurance; QC — quality control; SQC — statistical quality control; SPC — statistical process control; CIM — computer-integrated manufacturing.

Once the need for a LIMS has been justified, specifications have been written, and available vendors surveyed, the LIMS selection process begins. Each vendor's product must be evaluated for conformance to the specification and flexibility for future modifications. If an adequate solution is not available commercially, consideration must be given to constructing a custom LIMS solution. This could be done in-house or by contractors. If the custom option is preferred, resources needed for future modifications and maintenance should be factored into the cost.

Selection of LIMS software should take into account both data compatibility and ease of use. Since a LIMS is one information system within an organization, it may need to share data with other systems. Therefore, the software should be capable of sharing data directly or exporting data into compatible formats.

Database Management

A database management system (DBMS) is used by most LIMS systems for storing data. Examples of commercially available DBMS are DB2, DBASE, Informix, INGRES, ORACLE, and RDB. All of these DBMS conform to the "relational" model developed by Codd (19). Figure 3 demonstrates the use of a relational DBMS for storing LIMS data. Here data is grouped by type so customer and analysis requests are stored separately from sets of sample information which are, in turn, stored separately from sets of analysis results. Individual records are linked or related by unique identification data.

Table 4. LIMS Expenditures Based On Project Size^a

Laboratory feature	Small	Medium	Large
project cost, \$	60 × 10 ³	300 × 10 ³	1,500 × 10 ³
reports	20	70	300
people	5	50	500
laboratories	2	8	30
samples per year	6 × 10 ³	40 × 10 ³	300 × 10 ³
analyses per year	30 × 10 ³	250 × 10 ³	2,000 × 10 ³
analysis time, wk	1	3	15
expenditure, \$	3 × 10 ³	9 × 10 ³	45 × 10 ³

^a Ref. 15.

Relational databases can store unlimited numbers of results for every sample and unlimited samples for every request. The advantage of a relational DBMS over a more traditional hierarchical system, in which data sets may contain other data sets, is that the design of the database only has to consider relationships between data elements, not the number of instances for any given variable.

A DBMS performs what is called transaction management. This process allows multiple users to access and store data in the database without corruption. The ability to do this is particularly important when data are being written to the DBMS, because power interruptions or hardware failure can cause database transactions to be incompletely processed. Transaction managers use the "all or nothing" principle: all the data is written to the DBMS, ie, the transaction is completed, or none of it is written.

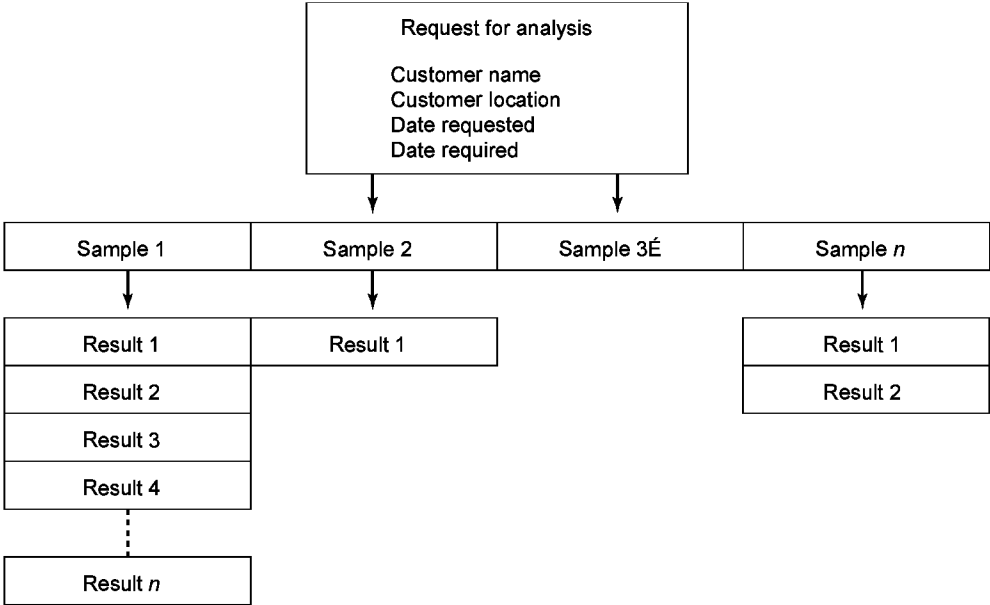


Fig. 3. Use of a relational database management system.

Another important feature of a LIMS DBMS is the ability to perform ad hoc database queries. It is impossible to predict all the forms in which LIMS users will want to display their data while the LIMS is being designed. As a result, it is desirable to select a LIMS which allows users to define their own reports. Most commercial DBMS have a standard query language (SQL) interface. SQL is a simple database query tool which is based on

```
English-language commands. This sample SQL query says, "Give me all sample request data containing John Smith as the customer."

SELECT* FROM SAMPLE REQUEST
WHERE CUSTOMER NAME   John Smith
```

Impacts of New Technology

Computer hardware costs have decreased dramatically. As a result, systems have become more affordable. Higher performance of new technology allows more functional capacity to be provided in a smaller, less costly machine.

Advances in network operating systems (NOS) provide database server and independent, simultaneous (distributed) data processing capabilities necessary to support LIMS functions throughout a network of computers. Previously this was possible only with a centralized processing model such as in traditional minicomputers and mainframes. The cooperative network model, also known as peer-to-peer processing, uses networks to distribute data and processing. This allows construction of systems that can be expanded as needed. The independence of each machine eliminates problems inherent in the centralized model, such as a single point of failure and high start-up and maintenance costs. Peer-to-peer networks were thought to be easier to maintain and operate. This is not actually true, because the distribution of data required for independence forces systems to store data temporarily and forward it to a central location to allow access by other network nodes. As a result, not all data is available in one place at one time, making efficient administration difficult for all but very small, highly tailored systems.

Another significant impact of new technology is the evolution of the client server computing model into commercially viable systems. This model incorporates a more powerful computer for data storage and retrieval (the server), connected to client workstations via a network. Client computers perform processing for the user interface, and pass messages and data to the server. Clients are typically personal computers or other graphics-capable devices (such as X-Windows terminals). Under this model, processing can be distributed within a highly integrated, secure data storage environment. Client server architectures achieve a system which provides the independence of the cooperative network (peer-to-peer) computing model, but is highly scalable, secure, robust, and easily administered. It has the added advantage of presenting a graphical user interface (GUI) to the LIMS user, in which computers, file systems, and programs are represented by simple icons, providing a familiar environment while hiding network and server complexities.

A significant concept of the client server model is to extend the scope of the application to function in an enterprise-wide (possibly worldwide) network of interconnected LANs, which allow LIMS and other applications following the client server model to be operable and administrable on a much larger scale than either LAN or central processing models.

The client server model often allows easier integration with other network applications (eg, finance, project management, or human resources) which typically operate in the environment of the server component of the client server system. Client server can be gradually introduced in an existing minicomputer environment, often with little adverse incremental impact in terms of retraining and additional cost.

PC workstations have become powerful, are simpler to use, and are generally ubiquitous within both laboratory and office environments. Using Windows-based software now widely available on instrument data stations as a basis for the LIMS will reduce the training costs and capital requirements when implementing a LIMS.

Networks required for peer-to-peer LANs often already exist in some form as a result of other office integration efforts. The existence of these networks can often reduce LIMS implementation costs by taking advantage of the reusability of the systems and networks already present in the laboratory.

Databases are becoming more standardized, thereby allowing a greater number of supplemental functions to be added to a LIMS and foreign systems, eg, project accounting, to be more easily integrated.

For these reasons, the desktop and client server models are expected to increase in percentage of LIMS software offerings and installed base in the future.

Instrumentation advances have increased the power and quality of the fundamental analytical techniques used in conjunction with LIMS. Unfortunately, these advances come at a price of increasing complexity and volume of information. Despite all of the architectural and technological advances of computer hardware and software, the demands of the information requirements still exceed the computing capabilities, so as to put continuing pressure on computer manufacturers to increase storage and processing capabilities even further.

Modern analytical instruments are tending toward results which cannot be reduced to the single-valued results easily entered into LIMS. Many of the newer instruments produce extremely complex results in the form of tables, spectra, images, or multidimensional relationships which are not easily represented in databases using the relational model employed by most current LIMS. Existing choices range from using a nonnumeric completion characteristic, which might reference a secondary computer file containing the complex results (this file may or may not exist on the same computer as the LIMS), to an alternative which treats the complex result as single objects and shifts to the use of object-oriented database tools to achieve the desired information within the LIMS itself. Shorter-term integration with existing file and document management systems may be used to simulate the object-oriented database (OODB) concept.

Data acquisition has evolved, but standards are still lacking (20). This makes data acquisition the most difficult and time-consuming aspect of the overall LIMS implementation. The time savings which result from automated data capture result in its generally being undertaken, to some degree, despite the difficulties. Other than chromatography laboratories and others which have a more uniform instrumentation environment, the data acquisition portion of the LIMS implementation tends to be a custom integration development project. This results in a relatively fixed system which can be adversely affected by changes (upgrades) to underlying analytical subsystems.

Work is being done to create uniform standards for exchange of information between analytical instrumentation and external (host) computers, but the diversity and the competitive nature of the instrumentation marketplace tend to impede these efforts, leading to an environment of constant change and a need for new and rewritten programs to communicate between LIMS and the automated instruments.

BIBLIOGRAPHY

1.

Good Laboratory Practice for Nonclinical Laboratory Studies, 21 Code Fed. Reg., Part 58, U.S. Food and Drug Administration (FDA), U.S. Government Printing Office, Washington, D.C., Apr. 1984.

2.

L. Sattler, E. Lonens, and A. Krumm, *Amer. Lab.* **21**(9), 6 (1989).

3.

R. D. McDowall, ed., *Laboratory Information Management Systems: Concepts, Integration and Implementation*, Sigma Press, New York, 1987, p. 12.

4.

A. Braithwaite, *Anal. Proc.* **24**, 126 (Apr. 27, 1987).

5.

R. Keeler, *R&D Magazine*, 37 (Apr. 1991).

6.

R. D. McDowall, *Lab. Info. Mgt.* **17**, 270 (1992).

7.

R. R. Stein, *Process. Ctrl. Qual.* **2**, 3–14 (1990).

8.

J. R. Jordan, *Anal. Consumer* **3**(11), 10–15 (1993).

9.

R. T. Martinnot, *Today's Chemist at Work*, SR3–7 (Apr. 1992).

10.

E. L. Cooper, R. Hice, and P. D. Rahn, *Sci. Comput. Automat.*, 45–49 (Feb. 1992).

11.

R. R. Stein, *Lab. Info. Mgmt.* **13**, 15–36 (1991).

12.

Ibid., p. 21.

13.

Ref. 11, p. 22.

14.

Ref. 11, pp. 24–25.

15.

Ref. 11, p. 19.

16. T. J. Long, *Chemometr. Intell. Lab. Sys.* **17**(3), 289 (1992).
17. J. R. Jordon, *Anal. Consumer* **3**(11), 4 (1993).
18. Ref. 4, p. 125.
19. J. Martin, *Principles of Database Management*, Prentice-Hall, Inc., New York, 1976.
20. J. R. Jordon, *Anal. Consumer* **3**(11), 7 (1993).

Michael G. Barrett
Kenneth O. MacFadden
Jean A. Taylor
W.R. Grace & Co.-Connecticut



LACQUERS.

See COATINGS; PAINT

LACRIMATORS.

See CHEMICALS IN WAR.

LACTIC ACID.

See HYDROXYCARBOXYLIC ACIDS

LACTONITRILE.

See CYANOHYDRINS

LACTOSE.

See CARBOHYDRATES; MILK PRODUCTS; SUGARS

LAMINATED AND REINFORCED METALS.

See COMPOSITE MATERIALS; METAL MATRIX COMPOSITES

LAMINATED AND REINFORCED PLASTICS.

See LAMINATED MATERIALS, PLASTICS; REINFORCED PLASTICS

LAMINATED MATERIALS, GLASS

A laminate is an orderly layering and bonding of relatively thin materials. A commonly laminated material is glass. Most commonly, two pieces of float or sheet glass are bonded with poly(vinyl butyral) [9003-62-7] (PVB) (see VINYL POLYMERS, POLY(VINYL ACETALS)) to produce a highly transparent safety glass, eg, an automotive windshield. This combining of transparent abrasion-resistant glass and resilient plastic achieves the durability and safety demanded of such products. Other materials that may be incorporated in laminated glass are colorants, electrically conducting films or wires, and rigid plastics. The value of the laminate is the utilization of the desirable properties from each of the constituents. In the case of laminated glass, the excellent weathering properties of the glass protect the impact energy-absorbing plastic interlayer from deterioration, abrasion, and soiling.

Benedictus, a French chemist who accidentally broke a flask that contained dried-on cellulose nitrates, is credited with founding the laminated-glass industry (1). The first patent was issued in 1906 (2). The growth of the laminated glass market was slow until automobile numbers and automotive speeds increased to the point that glass-caused injury was of concern. By the late 1920s, laminated windshields were standard in automobiles. The most common construction was two pieces of plate glass bonded with cellulose nitrate. However, the plastic interlayer introduced problems of haze, discoloration, and loss of strength, and it was replaced by cellulose acetate in 1933. Cellulose acetate demonstrated improved stability to sunlight but lacked strength over a broad temperature range and produced haze. The advent of the poly(vinyl butyral) resins in 1933 permitted the development of the modern interlayers that are used to make the majority of laminated safety glass in use; the resins were adopted for all automotive laminates by 1939.

Laminated glass is not a true composite material (see COMPOSITE MATERIALS). The glass needs the safety net effect of the interlayer if impacted, and the interlayer needs the durability and rigidity of the glass for useful service other than during impacts. Exceptions where laminated glass more truly fits the definition of a composite are when it is used for noise attenuation (see INSULATION, ACOUSTIC) or bullet resistance. In these applications, the alternate layering of rigid and soft materials achieves results beyond those produced by either alone.

Properties

Laminated materials frequently have limits on properties below those found in one of the components. Laminated glass with a PVB interlayer has a maximum service temperature not exceeding 70°C, far below that of solid glass. The strength of laminated glass is dependent on the number, thickness, and strength of the individual glass plies and on the characteristics of the particular interlayers used. For the majority of laminates consisting of two plies of annealed glass and one PVB interlayer, the bending strength is about 0.6 of that for an equal thickness of solid glass.

Glass–PVB laminates become more rigid with a decrease in temperature, and below –7°C approach the performance of solid glass. At temperatures above 38°C these laminates are less rigid and provide improved penetration resistance. Some applications utilize heat-strengthened or tempered glass for

additional strength. Figure 1 is an example of a wind-load chart for the combination of heat-strengthened and laminated glass (3). Wind-load information is used jointly by the architect, glazing contractor, and glass manufacturer to determine the permissible glazing area and glass thickness required to meet the design wind load.

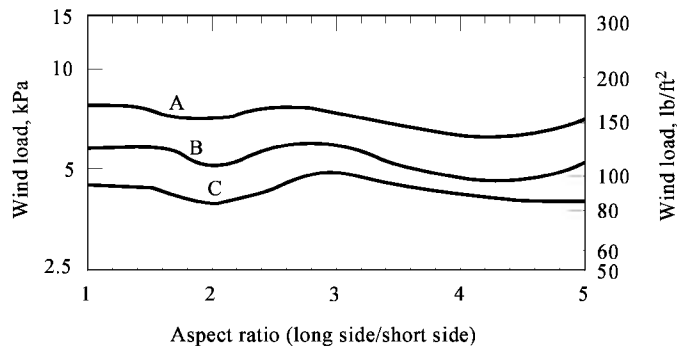


Fig. 1. Wind-load data for heat-strengthened and laminated 3.2-mm glass. Architect's specified probability of breakage is 8–1000 laminates for a 1-min uniform wind-load duration. Four sides supported in weathertight rabbet. Curves for different glazing areas: A, 0.93 m² (10 ft²); B, 1.39 m² (15 ft²); C, 1.86 m² (20 ft²).

Courtesy of PPG Industries, Inc.

Most laminated glass applications are concerned with impact strength, and minimum performance levels are required by specification. The impact strength of two plies of laminated, annealed glass and various PVB thicknesses are available (4). Aircraft laminates may utilize electrical resistance heating as deicing for vision enhancement.

Automotive and architectural laminates of PVB develop maximum impact strength near 20°C, as shown in Figure 2. This balance is obtained by the plasticizer-to-resin ratio and the molecular weight of the resins. It has been adjusted to this optimum temperature based on environmental conditions and automobile population at various ambient temperatures. The frequency and severity of vehicle occupant injuries vs temperature ranges at the accident location have been studied (5), and the results confirm the selection of the maximum performance temperature and decreasing penetration resistance at temperature extremes.

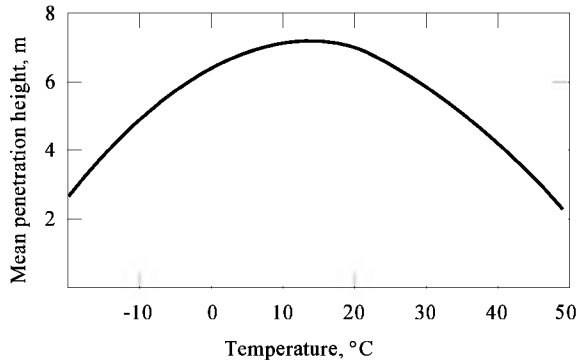


Fig. 2. Typical penetration resistance vs temperature data from laboratory procedure (305 × 305-mm laminates, 0.76-mm PVB; 2.27-kg ball impact).

Courtesy of Monsanto Co.

The optical properties of laminated glass are required to be equal to solid glass, because most applications are in vision areas. Light scattering by the interlayer essentially is nonexistent if PVB with the correct index of refraction is used. Clean-room practices can reduce the dust and lint that is attracted to the surfaces (see STERILIZATION TECHNIQUES). Visible-light transmittance of a typical automotive laminate (2.1-mm glass, 0.76-mm PVB, 2.1-mm glass) is nearly equal to solid glass of the same thickness (Fig. 3), and noticeable color change usually is absent. Visible-light transmittance is about 88% for clear glass laminates and ranges from 70 to 80% for windshields made with tinted glass. Sunroof laminates have been made with as little as 4% transmittance to reduce solar load. All uv light is absorbed below 370 nm and several discrete absorption bands are in the infrared beyond 1100 nm. The uv absorption may be enhanced when additional protection of color dyes is required, eg, gradient shade bands in automotive windshields or merchandise in window displays. Generally, the solar uv transmittance is on the order of 30–35%, and the infrared is about 97% for 0.76-m thick PVB.

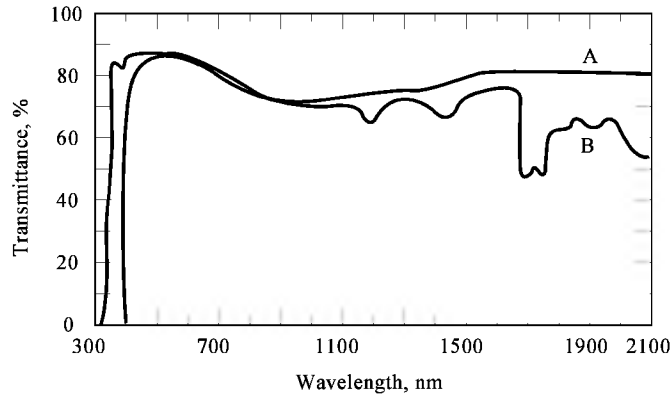


Fig. 3. Visible-light transmittance of automotive laminate: A, 475-mm monolithic glass; B, 5-mm laminated glass.

Courtesy of Ford Motor Co.

The index of refraction of PVB (1.48) is close enough to glass (1.520) to couple the two glass plies with a reflectance loss of only 0.02%.

$$R = \frac{(n_2 - n_1)^2}{(n_2 + n_1)^2} \quad (1)$$

where R is the reflectance at the interface and n_1 and n_2 are the different refractive indexes. The absorption coefficient for visible light (400–700 nm) generally is 0.25 to 0.45 for PVB. This produces a transmission loss within the PVB of 0.7–1.3%, which is mostly in the blue and ultraviolet portion of the spectrum.

Subsequent to the lamination process, some defects may appear that were not visible previously in the glass. One of these phenomena is called a bull's eye when found in windshields. These are small depressions that are formed in the glass pair during bending by the presence of glass chips or other debris between the plies. Upon lamination, the pockets fill with PVB and become convex lenses. Conversely, shallow ridges on an internal glass surface may be absorbed in the PVB and the optics are improved. Various other optical distortions may be caused by nonparallel plies of glass or PVB.

Manufacture

Practically all conventional laminated glass utilizes plasticized poly(vinyl butyral) (PVB) as the interlayer. Curved, laminated windshields are by far the principal products; silicone and cast-in-place urethane resins are sometimes used in specialty applications. Laminators purchase PVB in rolls up to 500-m long, up to 270-cm wide, and from 0.38- to 1.52-mm thick. There are several plasticizers used and at different ratios of plasticizer-to-resin content, depending on the product being manufactured. Flexol 3 GH, (bis(2-ethylbutanoic acid), triethylene glycol ester), manufactured by Union Carbide, is utilized at about 44 parts per 100 parts of resin. Other plasticizers (qv) used are di-*n*-hexyl adipate and dibutyl sebacate. There are at least five companies offering these products with manufacturing facilities in the United States, Japan, Belgium, Germany, and Mexico. Because the plastic is an adhesive material, it is shipped either with a dusting or parting agent on the surface or is refrigerated so it does not cohere. The refrigerated material is clean, moisture adjusted, and ready to laminate. The dusted material requires washing and moisture conditioning. Removal of the parting agent by warm water, followed by a chilled water rinse, adds about 0.2% water content to the plastic which must be compensated either by overdrying before washing, or drying after washing so as to achieve the desired 0.3–0.5% H₂O content. These steps are done more efficiently on the continuous roll before cutting. Moisture content of the PVB is extremely important because it has a direct effect on the adhesion characteristics of the glass-plastic surfaces.

The drying stage is carefully controlled to relax the sheeting of physical stresses as well as to adjust the moisture content. It consists of draping the sheeting over slat conveyors or driven rolls in a temperature- and humidity-controlled oven. The gradient band sunshade that appears in many windshields is either printed continuously on the interlayer roll at the PVB manufacturing plant or extruded into the sheet at the time of manufacture. To permit a more pleasing conformance to the curved glass, the banded interlayer may be preshaped which causes the extremities of the band to be more nearly parallel the horizon in the installed windshield. The shaping of the interlayer may be carried out on the continuous roll using a cone-shaped expander prior to cutting the blanks (6). The radius of curvature is preset, depending on the pattern of the particular windshield being manufactured. The interlayer then is cut to approximate laminate size and accumulated in low stacks (150 mm max) ready for assembly. Another method for shaping the interlayer involves warping in special ovens after the blanks are cut. The interlayer stacks must be stored in cooled, moisture-controlled rooms to control water absorption and blocking of the highly plasticized material.

The glass for laminating may be annealed, heat-strengthened, tempered, flat or curved, clear or colored. Thicknesses of 1.5–12 mm are used. For flat laminates, the glass is cut to size, edged and treated, if specified, washed, and delivered to the clean room by conveyor. The washing process, in addition to cleaning, can affect the interlayer bond. Common water hardness residues at invisible levels can reduce adhesive strength of PVB to glass. The desired level is achieved by controlling the hardness of the final rinse water and by removing the water by air stripping as opposed to evaporative drying. The glass is cooled during drying to prevent premature sticking when the interlayer is placed on the glass, thereby permitting easier positioning of the components.

In order to manufacture curved laminates, the glass is preshaped before laminating. This is usually accomplished by simultaneously bending a pair of glass templates which are cut to the shape of the finished windshield and separated by an inert powder to prevent fusing of the plies. The bending process is typically carried out on a peripheral support, metal fixture, or mold; the pair slowly travels through a lehr so that the glass sags to shape by the force of gravity. Glass temperatures of 600°C are required to achieve the shape, and the shaping is followed by annealing to reduce stress. Banded windshields usually are constructed with one or more pieces of tinted, heat-absorbing glass to enhance occupant comfort and to reduce air-conditioning load.

The clean room typically is operated at 18°C and 26% rh, which produces an equilibrium condition for the desired interlayer moisture level. The interlayer is placed on one piece of glass, with the gradient band, if present, carefully positioned above the designated eye position. The adjacent piece is superimposed, excess interlayer is trimmed, and this "sandwich" is conveyed from the room through a series of heaters and rolls that press the assembly together while expelling air. Temperature is increased stepwise to 90°C and pressures of 170–480 kPa (25–70 psi) are applied. Solid rubber rolls usually are used with flat laminates, and curved glass requires segmented rolls on a swivel frame (Fig. 4) (7,8). For more complex shapes, peripheral gaskets may be applied and the assembly may have the air evacuated (9), or the entire assembly may be placed in a bag and the air evacuated. The bag may or may not be removed prior to autoclaving, but when using an oil autoclave where the oil would damage one of the components, an oil-resistant Teflon or poly(vinyl alcohol) bag can be used (10). The tacked assembly is loaded onto racks for autoclaving, which may be either in an air or oil vessel capable of pressing the sandwiches at 1.38–1.72 MPa (200–250 psi) and 100–135°C for 30–45 min. Curved laminates and multiple laminates may require longer cycles to allow the polymer material to flow completely.

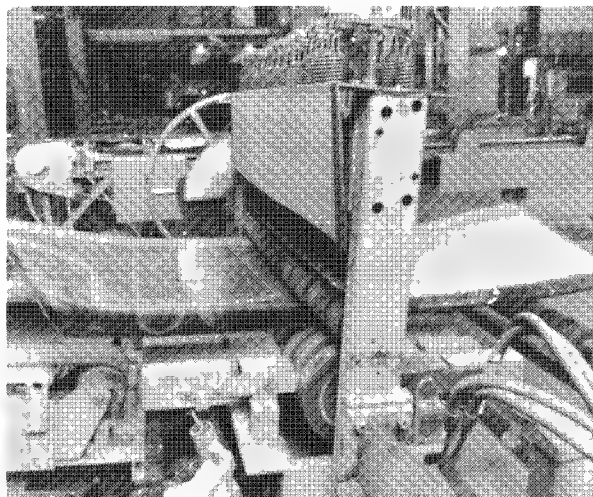


Fig. 4. The glass and vinyl "sandwich" is fed through a special de-air machine to remove trapped air and increase the adhesion of the vinyl to the glass.

Courtesy of Ford Motor Co.

In the autoclave cycle, the pressure is increased more rapidly than the temperature and then is maintained toward the end of the cycle as the temperature is lowered, to prohibit bubble formation and reduce any chance of delamination. The exit temperature must be no greater than 50°C to avoid thermal breakage. During the cycle, residual air is absorbed by the interlayer, and the embossed surface of the interlayer flows and wets the glass surface, thereby producing the clear laminate. Occasional small residual bubbles of trapped air can be removed by an additional autoclaving cycle. The air autoclave process is becoming the preferred method because it eliminates the subsequent washing of oily residues that are produced in oil autoclaving. The elimination of the process oil and subsequent washwater waste products also are environmental improvements. In some cases, additional trimming of the interlayer or finishing of the glass edge is required. The protruding interlayer may be trimmed or removed by wire brushing, but care must be taken to assure that the glass edge is not damaged in the process. The labeling of safety glass is done by grit-blasting through a mask or by silk-screening of a ceramic frit enamel. A final step in some windshield manufacture is the bonding of a small metal plate to the windshield. This plate, which is used to support the rearview mirror, is laminated to the glass using a special formulation of poly(vinyl butyral).

Radio antennas have been incorporated in some windshield models by inclusion of a very fine copper wire placed across the top of the windshield and vertically at the center. The wire is tacked to the interlayer prior to assembly and is embedded into the interlayer during autoclaving. An electrical connector that is soldered to the antenna wire is bonded to the bottom edge of the windshield for ease of connection. This type of construction is, however, being increasingly replaced by silk-screening of the antenna patterns directly on the glass using a silver-frit mixture.

Production and Shipment

Chemical attack, particularly from moisture and alkaline conditions, is prevented by use of acidic packing materials and open, ventilated packages. Good crate design and proper handling throughout shipment avoids mechanical damage. Glass-to-glass contact is never permitted. Long-term storage must be in well-ventilated areas, never in sealed containers. In the case of trans-ocean shipping, however, sealed containers are often used with a desiccant added to prevent moisture attack of the glass (see DESICCANTS).

Flat laminates are separated only by newsprint or plastic beads (ie, Lucite) and are bound into a block to prevent movement between the laminates. Curved laminates are spaced to prevent abrasion and require supporting dunnage at several points to prevent breakage by excessive flexing during shipment. Banding and blocking are designed to add compressive forces only. Staining is not a problem in the uncovered areas, but the supporting members are specified to have an acidic content to prevent chemical attack.

Laminated glass products are considered noncombustible and are shipped without DOT hazardous warning labels. Flat laminate packs, because of their high density, do not fill the car or trailer and require sturdy bracing to prevent shifting. Glass products are shipped and stored in a vertical plane, and during transportation they are placed so that each plate has an edge in the direction of travel.

Economic Aspects

The growth of laminated glass closely followed the growth of motor vehicles from the late 1920s to the 1960s. Windshields and side glasses of all domestic vehicles were laminated during this period. In the early 1960s, the flat, laminated, side glass was almost entirely replaced by curved, tempered glass. The curved windshield laminated glass market continued to follow automotive trends, but the flat glass products redeveloped around architectural uses. Architectural products currently represent 5–10% of the laminated glass volume. Increased consumer safety awareness and security needs are expanding the flat laminate market. Safety codes (eg, 16CFR 1201 and ANSI Z97.1) specify laminated glass as one means of meeting their requirements (11,12). In hurricane-prone areas laminated glass is being specified increasingly by local authorities.

Laminated windshields, as opposed to tempered glass windshields, are gaining in market share outside of North America. From 37% of the non-North American market of 1976, they were estimated to have reached 75% by 1982 (13). In addition to North America, Belgium, Italy, and the Scandinavian countries permit only laminated windshields, and other nations are increasing use by customer option. The trend toward laminated windshields is expected to continue and nonlaminated windshields will likely be obsolete by the year 2000 (14).

Specifications

Almost all of the laminated glass made is tested and certified to comply with certain safety performance standards. In the United States, there are two types of standards: automotive and architectural. For the former, *ANSI Z26.1-1973* is used and is incorporated in the Federal Motor Vehicle Safety Standard 205 (15). It specifies safety performance, durability, and optical quality. Specific tests are required depending on the location in the vehicle where the glazing is to be used. Item 1, the most difficult to meet, may be used in any location in the vehicle. It requires, in addition to other tests, support of a 2.3-kg ball dropped from 3.7 m onto a 305-mm square of laminate. Item 2 (safety glazing for any location except windshields) may be met by laminated glass using thinner PVB because it does not require the 2.3-kg ball test and the optical distortion tests. Laminated glass may also be used in locations specifying item 3 (no visible transmittance requirement) or item 11A (bullet-resistance glass; also requires appropriate tests, eg, ballistic tests). Automotive safety glass is required to be labeled as to the manufacturer, code, item, and model number that identifies the type of construction.

Conformance to the standard is achieved by submitting samples to an approved laboratory for evaluation and submitting the laboratory report to the American Automotive Manufacturers Association (AAMA). The approved certificate is sent to the manufacturers with copies to the state and provincial jurisdictions for which the AAMA serves as approvals agent (16).

Laminated glazing materials used in building locations specified by federal regulations are certified to comply to federal standards by the Safety Glazing Certification Council (SGCC) (17). Other locations requiring safety glazing specified by state or local code may use alternative standards (12). Glass complying to these standards is labeled permanently as to the standard (or standards) that it meets, including thickness, identification of the manufacturer, and plant. Also, it usually contains a date of manufacture. In situations where the large laminated sheets are cut into smaller pieces by the local distributor or installer, each piece is permanently labeled to indicate that it was cut from glass meeting the standard.

Certification to these standards is obtained by submitting a test report from an approved laboratory to the SGCC. Once certified, the product is assigned an SGCC certification number to identify it and the factory at which it was made. Subsequently, samples are selected randomly by the administrator at least twice a year to ensure continued adherence to the standard. Based on these re-evaluation reports, SGCC authorizes continued use of the certification label and the product listing published in its directory. The building standards are concerned mainly with body impact, and they require testing by impact on the glazing with a 45-kg bag. Detailed testing procedures and interpretations are available (11,12).

Bullet-resistant glass products are tested according to UL 752 (18). The test specifies that three shots are fired from 4.6 m and impacting within 100 mm of each other in a triangle, and that there is no penetration of the projectile nor any glass embedded in the corrugated board. The level of approval is determined by the velocity and energy level of the bullet at the muzzle of the firearm. Additional tests required include impacts 38 mm apart and tests over temperature ranges of 13–35°C for indoor use and –31.7 to 49°C for outdoor use.

The above-mentioned codes contain requirements for accelerated durability tests. In addition, interlayer manufacturers and laminators expose test samples for several years under extreme weather conditions, eg, the Florida coast and Arizona desert. The laminated products weather extremely well, with no change in the plastic interlayer. Occasionally, clouding is noted around the edges when exposed to high humidity for long periods, but this is reversible. Colored areas of PVB laminates may fade while subjected to extensive uv solar irradiation, which could cause an appearance issue. This has not, however, been shown to alter the laminate's other performance properties.

Analytical and Test Methods

Interlayer moisture is one of the important controls for PVB-to-glass adhesion of current formulations (although moisture-insensitive formulations are being developed). The moisture content equilibrates with the relative humidity to which the interlayer is exposed and thus is variable. Prior to lamination, interlayer moisture content is measured by one of three methods. The most rapid is by air absorption using a spectrophotometric technique to determine a

ratio of the 1925-nm to the 1705-nm wavelength peak (Fig. 3). A slower but less expensive method is weighing the interlayer before and after vacuum desiccation. The third and classical method is by Karl Fisher reagent; this technique is usually confined to instrument calibration exercises. The infrared method, in addition to being the most rapid, permits measurement of the interlayer moisture content while the interlayer is in the laminate. Instrumentation is available for monitoring interlayer moisture in full-size parts.

Interlayer bond strength is determined by either pummeling the laminate at 18°C to break away the glass and to determine the amount of adhering glass particles or by compressively shearing the laminate sample in a universal test machine. The optimum pummel range is three to five units on an arbitrary scale (from 1 to 10) established by the industry. The relationship of pummel value and compressive shear data to water content of the interlayer is given in Figure 5 and that of pummel value to mean penetration height is given in Figure 6. These data are influenced also by residual hardness of the water used to wash the glass.

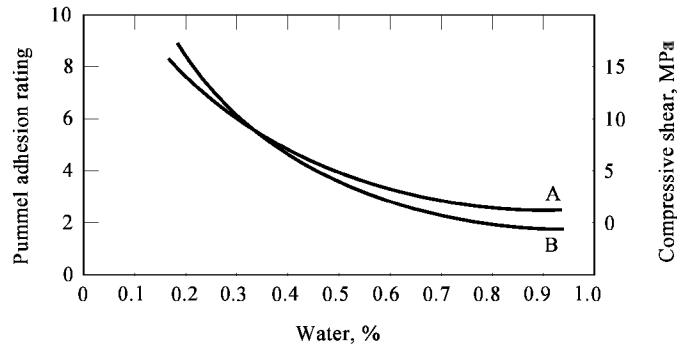


Fig. 5. Typical effect of moisture on PVB adhesion: A, pummel data (20°C) from Monsanto Co.; B, compressive shear data from Du Pont Co. To convert MPa to psi, multiply by 145.

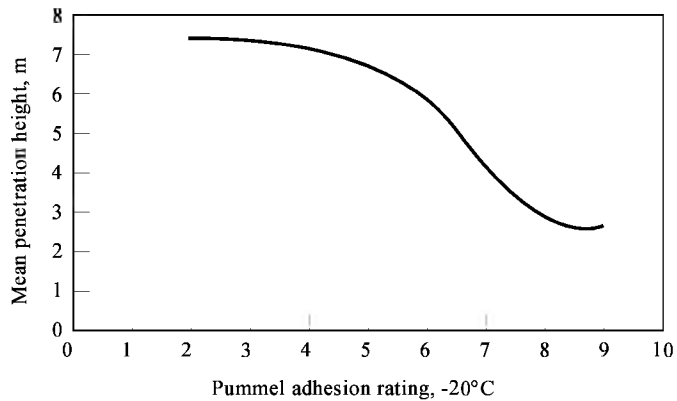


Fig. 6. Typical variation of mean penetration height with adhesion (2.27-kg ball, impact at 2°C on 305 × 305-mm laminates; 0.40% water, 0.76-mm PVB).

Courtesy of Monsanto Co.

Subsequent to processing, an inspection is made for incomplete bonding, inside dirt, and glass quality. In the case of windshields, rigid optical standards must be met, and these must be evaluated for the completed windshield. Extensive test requirements are described in the appropriate codes (11,12,15,18–24), and they include light stability, resistance to optical distortion, humidity, boil test, abrasion resistance, and assorted impact tests.

Uses

Penetration-Resistant Windshields. Performance difference between windshields manufactured in Germany and the United States was reported in the early 1960s (25,26). Three variables contribute to the greater safety of the German windshields against impact: thinner glass (especially the inboard member), thicker plastic interlayer, and higher moisture content of the interlayer. The latter acts as a plasticizer and in adhesion control (see PLASTICIZERS). By reducing the adhesion of the interlayer to the glass, more interlayer area can be released and stretched during impact. Also, the thicker interlayer, in addition to having more inherent strength, causes more fracturing of the glass during impact than in the older model of windshield. This, in turn, increases the amount of released interlayer for impact energy absorption. Upon review by the SAE Glazing Committee, it was agreed that the improvement was desirable, if it could be accomplished without taking the risk of increased water content. The U.S. PVB manufacturers subsequently developed controlled adhesion interlayers without increasing moisture above the previous standard content, and the glass fabricators utilized this material to produce laminated windshields with more than twice the impact resistance of the pre-1966 windshields. This product was introduced in limited production in 1965 and was used in all United States car lines for the 1966 models (27).

The ASA (now ANSI) performance code for Safety Glazing Materials was revised in 1966 to incorporate these improvements in windshield construction. The addition of test no. 26 requiring support of a 2.3-kg ball dropped from 3.7 m defined this level of improvement. It was based on a correlation established between 10-kg, instrumented, head-form impacts on windshields, on 0.6 × 0.9-m flat laminates, and the standard 0.3 × 0.3-m laminate with the 2.3-kg ball (28). Crash cases involving the two windshield interlayer types were matched for car impact speeds and were compared (29). The improved design produced fewer, less extensive, and less severe facial lacerations than those produced in the pre-1966 models.

Additional improvements have been incorporated since 1966 with the availability of thinner float glass. Glass thickness and interlayer thickness have been studied to optimize the product for occupant retention, occupant injury, and damage to the windshield from external sources (30,31). The thinner float glass windshields are more resistant to stone impacts than the early plate glass windshields. The majority of laminated windshields are made of two pieces of 2–2.5 mm annealed glass and 0.76 mm of controlled adhesion interlayer.

Special Laminated Windshields. Combinations of strengthened glass and interlayer offer advantages of lessened weight, higher impact resistance, lowered laceration potential, and resistance to bending stresses. These may be needed in high speed aircraft, helicopters, and motor vehicles. The additional strengthening can be achieved by chemical or thermal processes. The chemical process by ion exchange in molten potassium salts produces highly compressed skin and increases center tension. Thermal processes, capable of inducing high stress into float glass are also widely used, although not usually in automotive windshields.

Another variation of special construction (bilaminate) windshields consists of one ply of glass and one ply of an abrasion-resistant plastic. Although these laminate types have been available for several years, their limited mechanical durability has resulted in limited acceptance. Other features, such as deicing, defogging and solar rejection, can be incorporated into laminated glass. As technology advances it is anticipated that variable light transmission will also be possible by utilizing the properties of liquid crystal or electrochromic materials. These likely will need to be protected from the environment by

encapsulation in laminated glass.

Other automotive uses of laminated glass include colored glass and decorated glass. The privacy glass used in the side and rear glazing of vans can be made by laminating one or more layers of highly tinted PVB and clear glass. Opera windows containing metallic ornaments and sufficient plastic interlayer to accommodate their thickness have also been used. Laminated roof glazing can consist of a combination of coated glass and a colored PVB.

Aircraft Windshields. Aircraft windshields have extreme requirements in service temperature and pressurization and they must be resilient against high velocity bird impact. In addition, they must offer excellent visibility, both from optics and deicing capabilities, and an aerodynamic design. These highly specialized windshields are produced in low volume and are made by few companies, eg, Sierracin, PPG Industries, and Triplex Safety Glass Co. Construction varies with the need and service potential of the aircraft. Small planes of limited altitude and speed usually have acrylic monolithic windshields treated with a hardcoat material such as polysiloxane. Slower commercial aircraft use flat laminated glass made with aircraft-grade PVB (Monsanto Saflex PT). These aircraft require deicing capability which may be given by a conductive film that is pyrolytically or vacuum deposited on a glass surface, or conductive plastic film that is laminated in the sandwich. The third general class of aircraft windshield is for the modern, commercial, wide-body aircraft. These windshields become extremely complex, large in size, and expensive. A fourth type is for high speed, low flying military aircraft where birds, high skin temperature, and gunfire warrant extremely complex construction. The third and fourth types are multilayer constructions; typical examples are shown in Figures 7 and 8 (32,33).

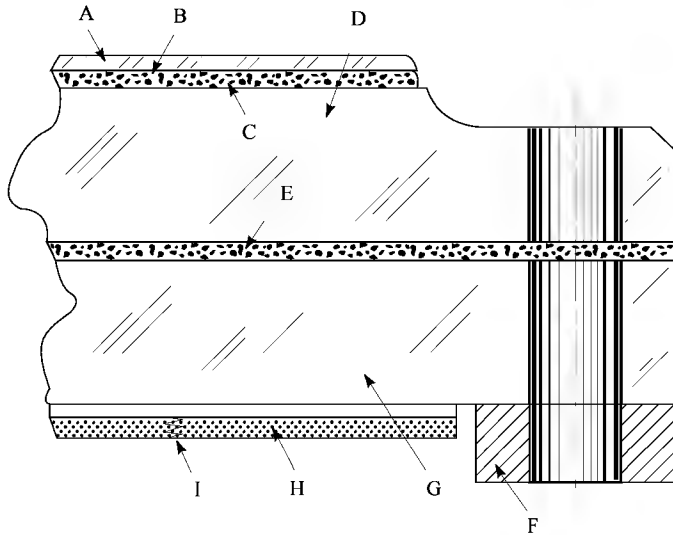


Fig. 7. Cross section of Sierracin windshield used on Boeing 747 (32): A, 2.2-mm chemically strengthened glass; B, Sierracote 3 conductive coating; C, 1.9-mm PVB; D, 23-mm stretched acrylic; E, 1.3-mm PVB; F, laminated cloth spacer ring; G, 23-mm stretched acrylic; H, 0.6-mm PVB; and I, 3.0-mm Sierracin 900.

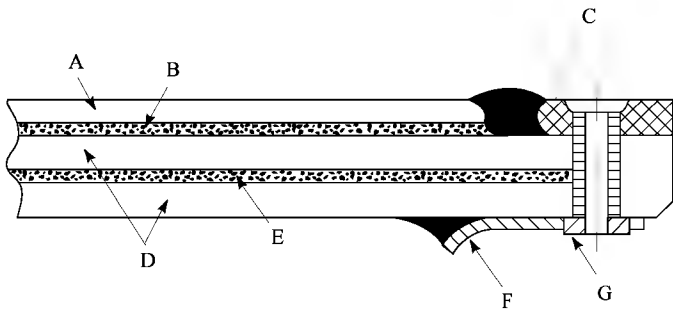


Fig. 8. Sierracin lightweight, birdproof F-111 windshield cross section (43). A, 3.0-mm as-cast acrylic face ply; B, S-100 silicone interlayer; C, fiber glass retainer; D, 6.4-mm polycarbonate structural ply; E, S-120 polyurethane interlayer; F, stainless steel bearing strip; and G, stainless steel bushing.

The Boeing 747 windshield (Fig. 7) is about 1.0 × 1.1 m and is curved to increase the pilot viewing area and to reduce air drag and air noise. Composed of seven plies, it weighs about 64 kg (32). The outer strengthened glass skin and the inner plastic shield may be replaced when damaged. The Triplex Safety Glass Co. also makes wide-body aircraft windshields, flat and curved, for Boeing and others. For the Boeing 747, two precurved, 12-mm plies of Ten-Twenty glass are laminated with PVB and covered with a 3-mm ply of Ten-Twenty glass bent to conform to the curved windshield. An electrically conductive coating, Hyviz, is applied to the inner surface of the outer ply and then is laminated to the 12-mm Ten-Twenty ply (34).

The construction of the F-111 windshield shown in Figure 8 replaced a glass–silicone laminate previously used. The all-plastic windshield has improved impact resistance so that it is birdproof to 250 m s (33). In this instance, the scratch resistance of glass was waived to obtain the impact performance at the allowed weight.

Architectural Products. Many specialized laminated glasses are made for architectural needs such as safety, sound attenuation, solar control, and security. These products may be further enhanced with colors and patterns for decorative effects. Safety glasses are specified in potentially higher risk breakage areas and overhead or sloped glazing (defined as more than 15° from vertical). Overhead glazing materials have varied in the past but more localities are accepting laminates. Sloped and overhead glazing frequently have heat-strengthened or tempered glass used in the construction of the laminate (35). Vertical passageway glazing usually is a 0.76-mm interlayer and sloped glazing is constructed with a 1.52-mm interlayer to accommodate the waviness of heat-treated glasses when they are used.

Noise attenuation is achieved effectively with laminated glass by the combination of the vibration damping effect of the plastic interlayer, an air gap, and usually an unbalanced glass thickness. Typical construction is 0.76–1.5-mm interlayer laminated with 3–10-mm glass. The type of glass or strength is not a factor in noise attenuation but a dead air space can be particularly effective in reducing selected frequencies. A Sound Transmission Class Index of 34–41 is achieved with single laminate glazing and can be improved if combined with double glazing that has large air spaces. Mounting of the glass in an airtight but flexible gasket reduces sound transmission (36). Airports, hotels, factory offices, and control rooms benefit from laminated acoustical glazing (37) (see INSULATION, ACOUSTIC).

Laminated glass is used for solar control, particularly where a highly reflective surface is not desired and where the laminate contributes other benefits. In these applications, a uniformly pigmented interlayer is obtained from the manufacturer and the laminate can be prepared by the conventional process. Broad ranges of colors and transmission levels are available with shading coefficients as low as 0.41. Pigmented interlayer is considered to be more

color stable than dyed interlayer. Browns, blues, greens, pink, white, and clear plastics containing uv absorbers are readily available. Body-colored glasses may be used also, usually with clear interlayer. In these cases, the laminate is dependent only upon the solar properties of the glass. Laminated glass may also be made using any of the low emittance (low-e) glass products on the market, but the low-e coating must not be in contact with the interlayer for the low emittance property to be achieved.

All laminated glass increases the level of security to some extent. However, depending on the application, security glass is constructed of multiple layers of glass, PVB, polycarbonate, polyurethane, or other polymer materials. Laminated glass permits the same visual observation as normal glass but prevents or delays unauthorized entry (or exit) until the attempt can be detected. It complies with test UL 972 (38).

Bullet-resistant glass is constructed of many layers of glass and aircraft-type PVB depending on the level of resistance desired. Typical products are 38–50 mm thick and weigh 90–130 kg m².

A third type of security glass is installed in modern penal institutions. This product is utilized for prisoner detention and obviates iron bars and their demeaning aspect. Typical construction utilizes three or more layers with at least one ply of thick PVB. Other constructions utilize polycarbonates, polyurethanes, and modified acrylics. Strengthened glass and electrically conductive circuits for alarms may be included. Large, heavy sections of similar construction have been used for underwater windows for boats, submarines, and aquariums. Four plies of fully tempered, 10-mm glass plus three plies of ~1.9-mm PVB totaling 44.5 mm in thickness has a modulus of rupture of 172 MPa (25,000 psi) (39).

Glazing of laminated architectural glass requires additional care in the selection of sealants and drainage design. Sealants (qv) must be free of solvents (particularly aromatics) and mineral or vegetable oils (3) and must not provide pockets that would trap water at the glass–PVB edge. Similarly, the glazing detail must be designed with proper drainage (35). Generally, the practice is similar to that of glazing organically sealed insulating units (40,41).

In the 1990s increased attention has been placed on the design and evaluation of transparent, architectural glazing panels that offer protection from sustained wind and snow loads, as well as gusts of hurricane strength winds (42,43). Research at academic institutions has been aimed toward improving window performance during extreme weather conditions. Many of the recommended penetration-resistant construction types consist of glass–plastic laminates (44,45).

BIBLIOGRAPHY

"Laminated Materials, Glass" in *ECT* 3rd ed., Vol. 13, pp. 978–993, R. M. Sowers, Ford Motor Co.

1. A. F. Randolph, *Mod. Plast.* **18**(10), 31, 98 (1941).
2. **U.S. Pat. 830,398** (Sept. 4, 1906), J. C. Wood.
3. *PPG Glass Thickness Recommendations to Meet Architects' Specified 1-Minute Wind Load*, PPG Industries, Pittsburgh, Pa., 1979.
4. R. G. Reiser and G. E. Michaels, *Proceedings of the Ninth Stapp Car Crash Conference*, University of Minnesota, 1965, pp. 181–203.
5. R. L. Morrison, "Influence of Ambient Temperature on Impact Performance of HPR Windshields," presented at *Fifteenth Stapp Car Crash Conference*, *SAE* 1971, pp. 603–612.
6. **U.S. Pat. 3,885,899** (May 27, 1975), D. J. Gurta and G. A. Koss (to Ford Motor Co.).
7. **U.S. Pat. 2,983,635** (May 9, 1961), R. E. Richardson (to Pittsburgh Plate Glass Co.).
8. **U.S. Pat. 3,009,850** (Nov. 21, 1961), J. P. Kopski and L. H. Schmidt (to Ford Motor Co.).
9. **U.S. Pat. 2,994,629** (Aug. 1, 1961), R. E. Richardson (to Pittsburgh Plate Glass Co.).
10. **U.S. Pat. 2,374,040** (Apr. 17, 1945), J. D. Ryan (to Libbey-Owens-Ford Glass Co.).
11. *Standard 16 CFR 1201*, Consumer Products Safety Commission, Bethesda, Md.
12. *Safety Performance Specifications and Methods of Test for Safety Glazing Material Used in Buildings*, *ANSI Z97.1-1975*, American National Standards Institute, New York, 1975.
13. R. C. Cunningham, *U.S. Glass Metal and Glazing*, U.S. Glass Publications, Memphis, Tenn., Jan. 1979, p. 28.
14. *Ward's Automotive Yearbook*, 39th and 41st ed., Ward's Communications, Inc., Detroit, Mich., 1977 and 1979.
15. *Safety Code for Safety Glazing Materials for Glazing Motor Vehicles Operating on Land Highways*, *Z26.1-1973*, American National Standards Institute, New York.
16. *Manufacturer's Guide for Safety Equipment Services*, American Association of Motor Vehicle Administrators, Washington, D.C., 1979.
17. *CPSC Certified Products Directory*, Safety Glazing Certification Council, Hialeah, Fla., 1980.
18. *Standard for Bullet Resisting Equipment* UL 752, *ANSI SE 4.6-1973*, Underwriters' Laboratories, Inc., Melville, N.Y., 1973.
19. *ASS As-R1-1968*, Standards Association of Australia, North Sydney, Australia, 1968.
20. *Brazilian Contran Resolution*, 483 74, Federal Official Gazette, Brazilia, Brazil, 1974.
21. *BS 5282-1975*, British Standards Institute, London, 1975.
22. *Specifications Relating to Safety Glass Requirements for Land Vehicles and Their Trailers*, Ministere De L'Equipement, Paris, 1975.
23. *Requirements on Safety Glass for Automotive Glazing*, Bundesministerium Ur Verkehr, Godesberg, Germany, 1973.
24. *A Tutte Gly Impettorati-Compartmentali Della Motorizzazione-Civile E Dei Trasporti N Concessione E Sezioni*, Ministero Dei Trasporti, Rome, Italy, Articles 218 and 297–302, 1959.
25. G. Rodloff, *Automobiltech. Z. (ATZ)* **64**(6), 1979 (1962); Eng. trans. 62-18916, National Translation Center, Chicago, Ill.
26. G. Rodloff, *Automobiltech. Z. (ATZ)* **66**(12), 353 (1964); Eng. trans. 62-11982, National Translation Center, Chicago, Ill.
27. J. C. Widman, *Recent Developments in Penetration Resistance of Windshield Glass*, *SAE* 650474, SAE, 1965.
28. E. R. Smith, presented at *Ninth Stapp Conference*, SAE, 1965, pp. 277–281.
29. D. F. Huelke, W. G. Grabb, and R. O. Dingman, *Automobile Occupant Injuries from Striking the Windshield*, Report No. Bio-5, Highway Safety Research Institute, Ann Arbor, Mich., 1967.
30. R. G. Rieser and J. Chabel, *Safety Performance of Laminated Glass Structures*, *SAE* 700481, SAE, 1970.
31. H. M. Alexander, P. T. Mattimoe, and J. J. Hofmann, *An Improved Windshield*, *SAE* 700482, SAE, 1970.
32. G. L. Wiser, "Sierracin® Glass Plastic Composite Windshields," presented at *Conference on Transparent Materials for Aerospace Enclosures*, U.S. Air Force and University of Dayton, June 25, 1969.
33. J. B. Olson, "Design, Development and Testing of a Lightweight Bird-Proof Cockpit Enclosure for the F-111," presented at *The Conference on Aerospace Transparent Materials and Enclosures*, Long Beach, Calif., Apr. 24–28, 1977.
34. R. W. Wright, "High Strength Glass in Service—A Status Report," presented at *The Conference on Aerospace Transport Materials and Enclosures*, Tech. Report AFML-TR-76-54, Atlanta, Ga., 1975.
35. *Archit. Rec.* (6), 143 (1979).
36. *Architectural Saflex® for Sound Control*, Tech. Bulletin No. 6295, Monsanto Polymers and Petrochemicals, St. Louis, Mo., 1972.
37. J. M. Clinch, *Study of Reduction of Glare, Reflection Heat and Noise Transfer in Air Traffic Control Tower Cab Glass*, *FAA-RD-72-65*, *AD747069*, NTIS, Springfield, Va., 1972.
38. *Burglary-Resisting Glazing* UL 972, Underwriters' Laboratories, Inc., Melville, N.Y., 1978.
39. *The New Look—Prisons Without Bars*, *Sierracin Field Report*, Sierracin Corp., Sylmar, Calif., 1972.
40. *Alum. Curtain Walls* **6**, 24 (Sept. 1972).
41. *FGMA Glazing Manual*, Flat Glass Marketing Assoc., Topeka, Kans., 1974.
42. W. L. Beason and J. R. Morgan, *J. Struct. Eng.* **111**(2) (1984).
43. R. A. Behr and co-workers, *J. Struct. Eng.* **111**(5) (1985).

44. R. A. Behr and co-workers, *J. Struct. Safety* **11**(1) (1991).
45. *Standard Practice for Determining the Minimum Thickness and Type of Glass Required to Resist a Specified Load*, ASTM Standard E1300-89.

General References

R. N. Pierce and W. R. Blackstone, *Impact Capability of Safety Glazing Materials*, PB195040, Southwest Research Institute, San Antonio, Tex., 1970; contains detailed descriptions of test equipment, methods, and results for all types of glazings.
SAE Transactions (annual), *SAE Handbook* (annual), Society of Automotive Engineers, Warrendale, Pa.
Stapp Car Crash Conference series (annual, 1956 on), Society of Automotive Engineers, Warrendale, Pa.; for safety and construction of automotive glass.

R. Terrell Nichols
Ford Motor Company

Robert M. Sowers
Consultant, Ford Motor Company



LAMINATED MATERIALS, PLASTIC

Laminates are materials made up of plies or laminae stacked up like a deck of cards and bonded together. Plywood is a common example of a laminate. It is made up of thin plies of wood veneer bonded together with various glues. Laminates are a form of composite material, ie, they are constructed from a continuous matrix and a reinforcing material (1) (see also REINFORCED PLASTICS).

High performance composites may be laminates wherein veils of carbon fiber are treated with an epoxy resin, stacked up to the desired final product thickness, and then laminated together under heat and pressure (see COMPOSITE MATERIALS; CARBON AND GRAPHITE FIBERS). Simply mixing together carbon or glass fibers and polymeric resins to form a reinforced plastic leads to a composite material, but this is not a laminate if not constructed from discrete plies.

Laminates are a special form of composite material or reinforced plastic because the continuous reinforcing ply of fibrous material imparts significant strength in the x - y plane. The strength along the z axis results from interlaminar bonding of resins. Very few fibers are oriented in the z direction, so it tends to be the weak link in this type of composite.

The reinforcing ply of laminates may be a woven fabric scrim, a nonwoven web of polymer monofilaments, or a mat of fibers. One of the most common reinforcements in use is also one of the oldest, ordinary cellulose fiber paper.

During the papermaking process, some degree of fiber orientation results in the machine direction. Typical papers are twice as strong in the length direction as in the cross direction. Therefore, because laminates have evolved into building-sized modules such as 4×8 sheets, all plies in the paper-based laminates are oriented in the same direction, and the laminate properties are quite anisotropic. The length or machine direction is much stronger than the cross machine direction, and there is even less strength between plies where there is virtually zero reinforcement.

As the laminate industry grew, this anisotropic behavior was accepted and fabrication techniques adapted to it. For example, expansion and contraction space was left between wall panels, very strong adhesives were developed for bonding the product to substrates, special substrates were qualified, and where it was necessary to cut holes into the laminates the corners were radiused to prevent cracking from stress concentration.

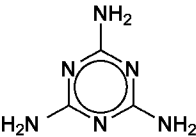
Early laminates tended to be small because available presses were small, and their original uses were to replace small parts such as the natural mica insulator boards used in radio chasses. As decorative laminates evolved from industrial laminates and the size grew to serve markets such as tabletops, countertops, and wall paneling, laminate dimensions tended to fall into the typical building module ratio of about two length to one cross, such as 2×4 s, 4×8 s, etc.

By comparison, high performance composite laminates are not only crossplied like plywood, but actually have laminae stacked at very specific angles to one another to achieve optimal uniform properties in the x - y plane (2).

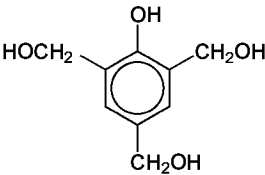
Resins

The commonly used resins in the manufacture of decorative and industrial laminates are thermosetting materials. Thermosets are polymers that form cross-linked networks during processing. These three-dimensional molecules are of essentially infinite size. Theoretically, the entire cured piece could be one giant molecule. The types of thermosets commonly used in laminates are phenolics, amino resins (melamines), polyesters, and epoxies.

The chemistry of melamines and phenolics is quite similar. In both cases formaldehyde [50-00-0] is added to the reactive sites on the parent ring to form methylol phenols (3) or methylol melamines (4) (see PHENOL RESINS; AMINO RESINS). There are six reactive sites on the triazine ring of melamine [108-78-1] (1) so it is possible to form hexamethylolmelamine. However, the most common degree of methylation is 1.5–2.0. The ortho and para positions of phenol are active; thus phenol can be trimethylolated (2). However, as with melamine, lower degrees of methylation such as 1.2–2.5 are common.



(1)



(2)

It should be possible to form linear noncross-linked polymers of melamine–formaldehyde or phenol–formaldehyde by reaction of one mole of the parent with one mole of formaldehyde, but this is generally not the case. The melamine crystal itself is very insoluble in water and only becomes soluble as the formaldehyde molecules add on. If much less than 1.5 moles of formaldehyde per mole of melamine are used, the aqueous resin solution is very unstable.

In the case of phenolics, it is possible to make linear thermoplastic polymers called novolaks, but this is done by reaction of less than one mole of formaldehyde with one mole of phenol; the resulting resin has a large excess of free phenol. Usually in application hexamethylene tetramine (HEXA) is added to the novolak. When heated, the HEXA breaks down into ammonia and formaldehyde and enters the reaction to form a light degree of cross-links in the final product.

Both melamine and phenolic resins usually exist as aqueous solutions although alcohols may also be used as solvents. In their A stage after the reaction of phenol or melamine with formaldehyde, these resins are not really polymers. They are simply methylolated melamines or phenols with a small quantity of dimers or trimers resulting from condensation of a methylol group with another reactive site. Monomers are also present. After saturating paper or another reinforcing material with these resins and heating to remove the water or other solvent, the resins are in a dry, solid state referred to as the B stage. Very little polymerization occurs because the treating temperatures do not reach much more than 100°C and then for only a very short time. In order

to condense into thermoset polymers, the C stage, much higher temperatures and longer times are required. The dried resins first melt and flow when heated in the press and eventually cure or thermoset as the temperature increases to the 130–150°C range for 20–40 minutes. Special fast curing resins can be cured at 170–180°C in about a minute or less.

The condensation reaction of methylolated melamine results in a methylene bridge linking two of the triazine rings. It is also possible for two methylol groups to condense together forming an ether linkage between two triazine rings. Considering that each triazine ring has an average of 1.5 methylol groups attached to it and a total of 4.5 remaining reactive sites that can condense with any methylol group to form a cross-link, it is obvious that a complex network structure results upon full cure. Modified melamines are available wherein one of the amino groups on the triazine ring has been substituted with a methyl group (acetoguanamine) or a phenyl group (benzoguanamine [91-76-9]). With less reactive sites, these polymers have lower final degrees of cross-linking. The situation with condensation of methylol phenols is exactly analogous.

Polyesters are also used in laminates, especially in low pressure laminate panels and flexible edge banding laminates or in special applications such as harsh chemical-resistant laminates. A typical polyester is made from one mole of isophthalic acid, one mole of maleic anhydride, and a slight stoichiometric excess of propylene glycol (5). The maleic anhydride provides the unsaturated sites for later cross-linking. Styrene or diallyl phthalate are the most common cross-linking agents. Unlike the melamines and phenolics that cross-link by condensation producing water as a by-product, the polyesters cross-link by free-radical addition at the double bonds (6). The free radicals may be produced from peroxides upon heating, from photoinitiators upon interaction with uv light, or they may be directly generated using high energy radiation such as that produced by an electron beam (see POLYESTERS, UNSATURATED).

The fourth common type of resin used particularly in industrial laminates is epoxy. Resins from the diglycidyl ether of bisphenol A are the most common type of epoxy resin. The ether is formed by reaction of bisphenol A with epichlorohydrin (7), and cross-linked to form a thermoset by opening of the oxirane ring. Epoxy resins (qv) can catalytically homopolymerize or form cross-linked heteropolymers by coreacting through the epoxide groups with a curing agent such as triethylene tetramine. Since these often complex curing agents become part of the polymer structure, they can strongly affect mechanical properties.

Reinforcements

The reinforcing ply in a laminate may make up half or more of its total weight. Therefore, properties of a laminate are strongly dependent on the ply. These plies are specified by basis weight in grams per square meter and may range from as low as 15 g m² for a lightweight overlay sheet to as much as 300 g m² or more for a strong filler sheet. The caliper of the ply varies with its basis weight and the amount it was calendered during its manufacture. Therefore, the reinforcing ply density and porosity vary widely. Low density papers are easier to saturate, but tend to weaken when wet. High density papers are stronger and better as print surfaces, but accept resin poorly.

The most commonly used reinforcement for high pressure decorative and industrial laminates is paper (qv). The strong substrate layers, or filler, are kraft paper. Kraft is a brown paper made from a sulfate pulp process (8). It consists of both short cellulose fibers from hardwoods and long fibers from conifers. The long fibers impart most of the wet strength required for resin saturation processes.

For use in decorative laminates, the surface paper must be highly refined pure cellulose (qv) technically called alpha cellulose. It is made by the sulfite process. The alpha cellulose content of a given paper is determined by the amount of material remaining undissolved in 18% sodium hydroxide. Pure cellulose paper is required because this decorative layer will be pigmented and sometimes printed with a pattern such as a wood grain. It must be colorless initially, and it must be stable to heat throughout processing. The laminate must also be light-stable throughout its useful life.

Decorative papers are pigmented for color and opacity. Titanium dioxide is commonly used for whites and in light pastel shades. The TiO₂ may constitute 35% of the total raw paper weight. Thus, pigmented papers perform very differently from nonpigmented ones, affecting both processing and final physical properties. This is because a 50% resin, 50% cellulose composite may actually be 50% resin, 35% cellulose, and 15% inorganic filler. Organic dyes and pigments are also found in decorative papers used in laminates. Being organic, they do not contribute to the ash content nor have much impact on physical properties. Sometimes rayons are used in the surface layer. Rayons are derived from further chemical reaction of already purified celluloses.

Other reinforcements that may be used in the substrate layers of decorative laminates and throughout the structure of industrial laminates are woven fabrics of glass or canvas and nonwoven fabrics of various polymeric monofilaments such as polyester, nylon, or carbon fibers. Woven and nonwoven fabrics tend to be much stronger than paper and have much more uniform strength throughout the x-y plane. They greatly enhance properties of laminates such as impact and tear strength.

The reinforcing ply acts as the carrier for the plastic resin during intermediate processing steps known as saturation and B-staging. It is this ply that together with the resin makes a laminate a composite material, and the layering of these plies that makes the final product a laminate.

Manufacture

Treating. Treating is the term used in the laminate industry for the application of the plastic resin to the reinforcing ply or carrier web that eventually forms a ply in the composite laminated product. Typical means to apply the resin are reverse roll coaters, dip and scrape, or dip and squeeze operations. In reverse roll coating, the resin is applied to only one side of the carrier web with the coating rollers metering on the correct amount of resin solution. For the dip and scrape or dip and squeeze methods, the paper is submersed beneath the resin solution, and then excess resin is scraped off with a doctor knife or the treated web is run through a set of metering rolls or squeeze rolls to control the pick-up. Although reverse roll coating is a more precise method, dip and squeeze is easier and the equipment less costly. Pick-up of resin is also affected by the porosity of the paper and the viscosity and solids of the resin solution. A more porous paper absorbs resin more rapidly and usually soaks up a greater quantity. This is particularly important in reverse roll coating when only one side of the paper is exposed to the resin. A less viscous resin penetrates better, but if viscosity is reduced by decreasing solids, final pick-up of dry resin may actually be reduced. Viscosity may also be reduced by using a resin of lower molecular weight or by use of a better solvent for the resin.

After the resin is applied to the paper, the wet treated web enters a drying oven where most of the solvent is evaporated off. Modern treaters, such as that shown in Figure 1, have air flotation ovens so that the web is never touched by hot conveyor bars that may cause streaks noticeable in the final product.

As the treated web passes through the oven, the solvent evaporates and the web temperature increases somewhat. However, due to the speed of the operation, normally 30–300 m min, dwell time in a typical 30–40 m long oven is very short. Since there is also evaporative cooling until most of the solvent is removed, the treated paper exiting the oven is not much hotter than the boiling point of the solvent. The oven drying process with thermosetting resins is referred to as B-staging. There is considered to be advancement of the resin during this process, but actually, due to the short time and low temperatures, nothing more than drying occurs.

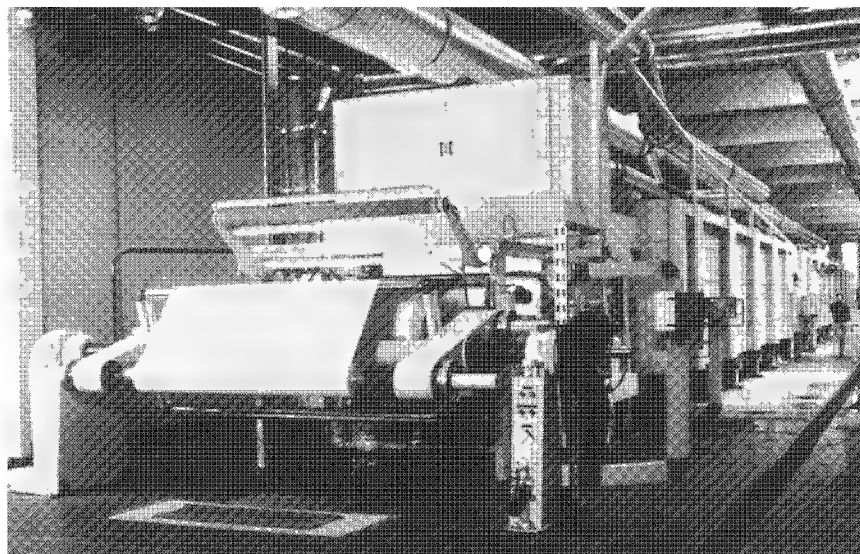


Fig. 1. Treating machine equipped with an air flotation oven.

Photo courtesy of VTTS Babcock Maschinenbau, GmbH.

The volatile content of the treated paper is important because moisture acts as a temporary plasticizer to promote resin flow during early stages of pressing (9). Dynamic mechanical analysis of the treated paper is a very useful means to study the initial flow stages of a resin and the cure time required to complete cross-linking (10).

By use of a three unit beta-gauge system on a treater it is possible to achieve excellent control of treating parameters. The first gauge traverses the raw paper and makes a continuous record of its basis weight. The second gauge is positioned after the coater head and measures the add-on weight of wet resin. The third gauge at the oven exit measures the dry resin weight plus any residual volatile. With feedback loops to the coater head, very precise control can be maintained. At the exit of the oven, the treated paper may either be sheeted or rerolled. If it is rerolled, it must be later sheeted in a separate operation prior to collation.

In the case of some types of polyesters, the dried treated sheet is very tacky, and must be interleaved with a plastic release film to prevent blocking in the roll prior to use. Certain polyesters and acrylics can be converted to a gelled state by exposure to actinic radiation.

Collation. Collation is the process by which the individual laminate plies are assembled prior to curing in the press. The buildup of the laminate determines the final properties of the product. The topmost sheet in the buildup may be a texturing or embossing paper as well as being a release sheet to allow for separation of the laminate from the caul plate used to mold it.

The texturing sheet may also contain a plastic coating that transfers to the laminate surface during the laminating process to provide special surface properties. The caul plate itself may be smooth or textured and act directly as the release. Immediately beneath this release sheet, there may be a decorative layer or a ply that imparts special properties such as a metallic foil for electrical conductivity or a static dissipating layer. If the decorative layer is a printed pattern, it may be covered by a special layer that transparentizes during the laminating process. This transparent ply provides a wear-resisting layer above the printed pattern.

Beneath the special surface layers, there are usually several plies of materials designed to provide physical strength to the laminate. In the cases of decorative and industrial laminates based on paper-reinforcing plies, they are normally stacked in the same direction with the length dimension parallel to the machine direction of the paper. This gives laminates noticeable directional properties. In high performance laminates used in aerodynamic applications, the plies are often crossed or even arranged at very specific angles to provide more isotropic properties.

When multiple laminates are pressed in a single opening between heated press platens (daylight), it is necessary to separate them. They are usually pressed face to face, back to back. A sheet of wax paper or plastic film can be used between the backs of two laminates to separate them. Smooth caul plates are inserted between two laminate faces. The plates are the tools used in flat-bed press processing and have an important function in molding the planar laminate surfaces as well as acting as release and texturing media. Laminates can be press cured in plateless (veneer) packs, but such laminates tend to have surface waviness resulting from nonuniformities in the multilayers of reinforcement material, and are not generally satisfactory for decorative purposes.

Press Curing. The laminate as an article of manufacture is prepared in a flat-bed press. Modern high pressure presses may be as large as 2 m × 5 m, and low pressure presses are as long as 7.5 m. Normally, high pressure presses have multiple openings or daylight, sometimes 20 or more, as shown in Figure 2. Low pressure presses have only a single daylight. The high pressure presses are designed to be operated up to pressures of 14 MPa (2000 psi) and the low pressure presses up to about 2 MPa (290 psi). Continuous presses are also in use for the manufacture of low pressure products. Some continuous high pressure presses are being built, but they are not in wide usage.

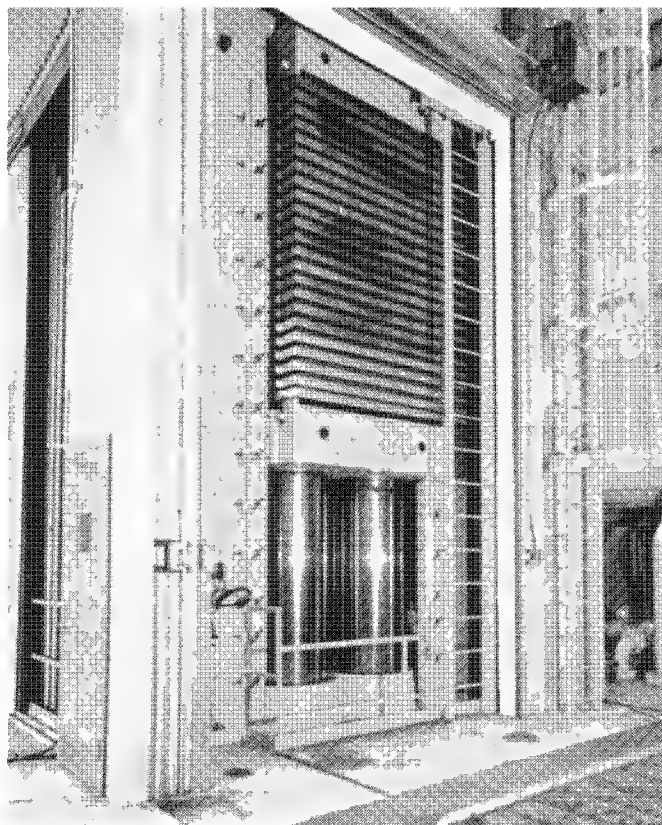


Fig. 2. Multiopening flat-bed laminating press.

Photo courtesy of G. Siempelkamp, GmbH & Co.

During the press operation, which is actually a form of compression molding, the resin-treated laminate plies are heated under pressure and the resins cured. The initial heating phases cause the resin to melt and flow into voids in the reinforcing ply and bond the individual plies together. The applied heat simultaneously causes the resin to polymerize and eventually to cross-link or gel. Therefore, resin viscosity reaches a minimum during the press cycle. This is the point at which the curing process becomes dominant over the melt flow process. Dynamic mechanical and dielectric analyses (11) are excellent tools for study of this behavior.

With a sufficiently long press cycle, a state of complete cure is reached. At this point, the laminate is cooled in the press, under pressure, and removed for finishing operations. If the press is opened at a temperature above the boiling point of trapped volatiles, vaporization occurs causing interlaminar blistering which ruins the laminate.

The quantity of resin applied to the reinforcing ply to achieve a state of full densification varies inversely with the laminating pressure. Therefore, high pressure laminates pressed at about 7 MPa (1000 psi) need only about 25–30% phenolic resin in kraft paper, whereas low pressure (1 MPa = 145 psi) laminates need 50–60% resin in the reinforcing ply if all voids are to be filled in the final product.

Low pressure processes are generally of very short duration (30–90 s), and high pressure cycles are typically one hour or more. Therefore, it is economical for the low pressure presses to be single opening or continuous whereas the high pressure presses must be multiple opening. Because of the short low pressure cycles and heat-transfer considerations, usually only one or two laminates or one panel can be pressed in the single daylight. By contrast, a dozen or more laminates may be pressed in a single high pressure press daylight, and these laminates may consist of 5–10 plies each. As laminates become thicker, the number per daylight must decrease because the opening is physically limited to about 50 mm by the overall press size and ram travel distance. High pressure laminates may be up to 25 mm thick, in which case only a single one is pressed in one daylight. The productivity of a large high pressure press in meters per minute may be very high in spite of it being a batch process. For example, 4-m long laminates pressed 12 per daylight in a 20 daylight press on a one-hour cure cycle are produced at a rate of 16 m/min. It is difficult for low pressure processes to match this rate of production and impossible for them to press multiple ply products because heat transfer never allows the interior plies to reach melt flow temperatures or the resins to cure during the short dwell times. Some short cycle high pressure presses are now entering the market. They offer increased manufacturing flexibility at about the same production rate as the older long cycle presses.

Low pressure presses are usually operated at about 180°C as compared to high pressure presses at about 135°C, but the extra low pressure press heat is not adequate to compensate for the short dwell time. Special fast curing resins must be used in conjunction with the low pressure presses.

Although flat-bed laminating is similar to compression molding, it is accomplished without edge constraints. In other words, the molds are open at the edges and theoretically the pressure in the x - y plane may fall to zero at the edge even though the z -direction pressure normal to the plane of the laminate is at 10 MPa (1450 psi). If it were not for the flow restraining characteristics of the paper ply normally found in high pressure laminates, it would not be possible to press in these open-edged molds. Edge effects do occur in laminates. Excess resin sometimes flows out at the edge and volatiles escape there. However, the extreme edges are normally trimmed away before the product is used.

Some high performance laminates consisting of carbon fiber webs and epoxy resins are cured in autoclaves. An autoclave is a pressure chamber in which the pressure is applied hydrostatically.

Finishing and Fabrication. Since laminates are normally pressure cured in flat-bed presses and plies overextend the plates, laminates have rough or uneven edges when removed from the press. These edges are sawed off and the back of the laminate is often sanded to improve the strength of subsequent bonding to various substrates.

Common grades of laminates tend to be thin materials ranging from 0.5–1.5 mm in thickness, therefore for most applications they must be supported. In the manufacture of furniture, cabinetry, and countertops the laminates are bonded to particle board or plywood. Since the laminates consist largely of cellulosic paper, their dimensional stability is similar to wood, particularly to particle board.

In small pieces or as inserts, laminates may be used unsupported because they are quite stiff and strong. The modulus of a high pressure decorative laminate is about 7 GPa (106^2 psi) at room temperature. Thick laminates range up to 25 mm and are very strong, having flexural strength of 130 MPa (19,000 psi). These products are used unsupported as toilet partitions, laboratory tops, and so forth.

An important fabrication operation for laminates is post-forming. This is an operation in which a laminate is heated and bent. It is really an anomaly because laminates are thermoset products as opposed to thermoplastics and as such are supposed to be cross-linked into an intractable network structure. However, in the early days of development of decorative laminates it was necessary to cut strips and bond a self-edge of the material to finish the appearance of a countertop. An obvious benefit would be to bend the laminate around an edge to finish it off. In order to accomplish this post-forming, special papers having a degree of stretch were used along with special resins. Also, press cure cycles specifically designed to limit the resin cure to a slightly

incomplete state were used. By doing so, it became possible to produce special laminates that could be rapidly heated to about 150°C and quickly bent to a simple radius of curvature. If the heatup is too slow, the laminate dries out and cure may advance with the end result being a crack in the bend. Because the paper-reinforcing ply in the laminate is not very extensible, complex bends cannot be made.

Properties and Grades

The properties of standard grades of high pressure decorative laminates manufactured are shown in Table 1. Tables 2 and 3 show properties of industrial laminates (12,13).

Table 1. Performance Properties and Values for Decorative Laminates

Type designation	Nominal thickness, mm	Thickness tolerance, mm	Wear resistance cycle ^a	Ball impact resistance, mm ^a	Dimensional change, ^a % ^{b,c}		Boiling water resistance rating ^{a,d}	Radiant heat resistance, s ^a
					MD	CD		
general-purpose, GP 50	1.270	±0.127	400	1270	0.5	0.9	NE	125
general-purpose, GP 20	0.508	±0.127	400	381	0.8	1.3	NE	60
post-forming, PF 42	1.067	±0.127	400	762	1.1	1.4	SL	100
specific-purpose, SP 125	3.175	±0.203	400	1905	0.3	0.7	NE	200
high wear, HW 120	3.048	±0.203	3000	1905	0.3	0.7	NE	200
fire-rated, FR 50	1.270	±0.127	400	1143	0.5	0.9	NE	75

^a Minimum.
^b MD = machine direction; CD = cross direction.
^c Maximum.
^d Rating system: NE = no effect; SL = slight effect.

Table 2. Performance Properties of 3.2-mm Thick Industrial Laminates^a

Grade	Flex strength, MPa ^b		Water absorption, % ^c	Permittivity, ^c 1 MHz	Dissipation factor, 1 MHz	Impact strength, J/m ^d		Dielectric breakdown parallel to laminations, kV
	Lengthwise	Crosswise				Lengthwise	Crosswise	
X	172	152	3.3	6.0	0.05	29.4	26.7	
XP	96.6	82.8	2.2					40.0
XPC	82.8	70	3.0					
XX	103	96.6	1.3			21.4	18.7	40.0
XXP	96.6	82.8	1.1	5.2	0.05			60.0
XXX	93.1	81.4	0.95	5.7	0.045	21.4	18.7	50.0
XXXXP and C	82.8	72.4	0.65	4.8	0.04			60.0
ES-1,2,3	93.1	93.1	1.8					
C	117	110	2.5			101	90.7	15.0
CE	114	96.6	1.6			85.4	74.7	35.0
L	114	100	1.6			72	58.7	15.0
LE	110	96.6	1.3	6.0	0.070	66.7	53.4	40.0
AA	124	110	2.5			192	160	
G-3	138	124	2.0			347	294	
G-5	307	262	2.0	8.0	0.080			23.0
G-7	138	124	0.35	4.2	0.022	347	294	32.0
G-9	379	241	0.70	7.4	0.018			60.0
G 10,11	379	310	0.15	5.4	0.035	374	294	45.0
FR 4,5								
N1	70	65.5	0.40	3.9	0.039	160	107	60.0
FR1	96.6	82.8	2.2	6.0	0.06			40.0
FR2	82.8	72.4	0.55	4.8	0.04			60.0
FR3	138	110	0.50	4.8	0.04			60.0
CEM-1 ^e	207	172	0.20	4.8	0.04			45.0
CEM-3 ^e	228	186	0.20	5.4	0.035			45.0

^a Values from NEMA LI-1-1089.
^b To convert MPa to psi, multiply by 145.
^c Permittivity = dielectric constant.
^d To convert J/m to ft-lb/in., divide by 53.38.
^e 2.4-mm product.

Table 3. Performance Properties^a of 3.2-mm Thick Industrial Laminates

Grade	Electrical strength perpendicular to lams, V/μm ^b	Insulation resistance, MΩ	Tensile strength, MPa, ^c		Modulus of elasticity, GPa ^d		Rockwell hardness, M	Density, g/mL
			Lengthwise	Crosswise	Lengthwise	Crosswise		

X	20		138	110	13.1	9.7	110	1.36
XP	19		82.8	62.1	8.3	6.2	95	1.33
XPC	17		72.5	58.6	11.0	5.5	75	1.34
XX	20	60	110	89.7	10.3	8.3	105	1.34
XXP	20	500	75.9	58.6	6.2	4.8	100	1.32
XXX	19	1000	103	82.8	9.0	6.9	110	1.32
XXXX, C	19	20,000 50,000	85.5	65.5	6.9	5.5	105 95	1.30
ES-1,2,3	22		82.8	58.6			118	1.40

^a Thermal expansion for all grades = 2.0×10^{-6} cm/cm °C; specific heat = 1.46–1.67 J (g·K) (0.35–0.4 cal (g·C°)); thermal conductivity 0.3 W (mK) (7.0×10^{-4} cal).

^b To convert V μm to V mil, multiply by 25.

^c To convert MPa to psi, multiply by 145.

^d To convert GPa to psi, multiply by 145,000.

Aesthetic properties are of greatest concern in decorative laminates. These include gloss, appearance, cleanability, wear resistance, stain resistance, and other surface properties. Physical properties are of most importance for industrial laminates. These include strength, electrical and thermal properties, expansion coefficient, and punchability. The definitions of the laminate grades in these standards follow.

DECORATIVE LAMINATES

General-purpose type is a high pressure decorative laminate (HPDL) designed for both horizontal and vertical applications where appearance, durability, resistance to stains, and resistance to heat up to 135°C (275°F) are required.

Post-forming type is an HPDL similar to the general-purpose type but is capable of being thermoformed under controlled temperature and pressure in accordance with the laminate manufacturer's recommendations.

- Cabinet liner type* is an HPDL intended only for use in cabinet interiors.
- Backer type* is an HPDL without a decorative face intended for use as a balancing sheet in panel construction.
- Specific-purpose type* is a general-purpose HPDL with increased thickness.
- High wear type* is also a general-purpose HPDL with increased surface wear resistance.

Fire-rated type laminate is an HPDL capable of providing fire-rated characteristics as determined by the test methods required by the authority having jurisdiction.

Miscellaneous Types. Various decorative effects have been developed which meet specific aesthetic requirements. These laminates may have special visual appeal, such as gloss finish, deeply embossed textures, and metallic surfaces. They are designed for specific installations and may not be suitable for all applications. For this reason, they are not included in these standards. Information concerning their proper application, properties, and care should be requested from the manufacturer.

INDUSTRIAL LAMINATES

Paper-Base Grades. *Grade X* is primarily intended for mechanical applications where electrical properties are of secondary importance. It should be used with discretion when high humidity conditions are encountered and it is not equal to fabric-base grades in impact strength.

Grade XP is primarily intended for hot punching. With good punching practice and depending on the complexity of the part design, sheets up to and including 1.6 mm (0.062 in.) may be punched at room temperature.

Grade XPC is primarily intended for cold punching and shearing. It is more flexible and shows higher cold flow but is lower in flexural strength than Grade XP.

Grade XX is suitable for usual electrical applications. It is not recommended for punching but may be machined using present state-of-the-art machine tools.

Grade XXP is better than Grade XX in electrical and moisture-resisting properties and more suitable for hot punching. It is intermediate between Grades XP and XXXP in punching and cold flow characteristics.

Grade XXX demonstrates good electrical properties in high humidity applications and has minimum cold flow characteristics.

Grade XXXP is better in electrical properties than Grade XXX and more suitable for hot punching. It is intermediate between Grades XXP and XXXPC in punching characteristics. This grade is recommended for applications requiring high insulation resistance and low dielectric losses under severe humidity conditions.

Grade XXXPC is similar in electrical properties to Grade XXXP and suitable for punching at lower temperatures than Grade XXXP. This grade is recommended for applications requiring high insulation resistance and low dielectric losses under severe humidity conditions.

Grade ES-1 is suitable for engraving as nameplates, etc. It is made with black or various colored surfaces and a white opaque core (usually melamine binder).

Grade ES-2 is similar in application to Grade ES-1 but made with a white subcore and black core (usually a phenolic binder) to obtain toughness when made in thick sheets.

Grade ES-3 is similar in application to Grade ES-1 but made with a white or various colored surface and black core.

Fabric-Based Grades. *Grade C* is made from cotton fabric weighing over 140 g m² (4 oz yd²). The maximum thread count in any ply is 28 cm (72 in.) in the fill direction, and the maximum total thread count in the warp and fill directions is 56 cm (140 in.). Heavier fabrics provide higher impact strength but rougher machined edges. Its use for electrical applications is not recommended.

Grade CE is made from cotton fabric with the same weight and thread count limits as Grade C. Suitable for applications requiring good moisture resistance, greater strength than Grade LE, and electrical properties nearly equivalent to Grade XX. This grade is not recommended for primary insulation for electrical applications involving commercial power frequencies at voltages in excess of 600 volts.

Grade L is made from fine-weave cotton fabric weighing 140 g m² or less. The minimum thread count in any ply is 28 cm in the fill direction, and the minimum total thread count in the warp and fill directions is 56 cm. Primarily suitable for machining and mechanical applications where finer edge detail and appearance than Grade C is required, Grade L has high density and good moisture resistance. Its use for electrical applications is not recommended.

Grade LE is made from cotton fabric with the same weight and thread count limits as Grade L. It is suitable for electrical applications requiring good moisture resistance and greater strength than Grade XX.

Asbestos-Based Grade. *Grade AA* is more resistant to heat than Grade C but not recommended for primary insulation at any voltage. It exhibits small dimensional changes when exposed to moisture.

Glass-Based Grades. *Grade G-3* is glass fabric with phenolic resin binder which shows high impact and flexural strength. It is used for thermal and mechanical applications and has good dimensional stability.

Grade G-5 is glass fabric with melamine resin binder and high mechanical strength. It is one of the hardest laminate grades, with good arc resistance; it meets UL94 V-0 when tested in accordance with UL94. This material has excellent electrical properties under dry conditions and good dimensional stability. Electrical applications should be limited to operating temperatures of 50°C (122°F) or less.

Grade G-7, glass fabric with silicone resin binder, shows extremely good dielectric loss properties under dry conditions and good electrical properties under humid conditions, although the percentage of range from dry to humid conditions is high. This material has excellent flame, heat, and arc resistance and meets UL94 V-0 when tested in accordance with UL94. It also has good impact and flexural strength.

Grade G-9, glass fabric with moisture-resistant melamine resin binder, is similar to Grade G-5 but with better electric strength properties under wet conditions. Electrical applications should be limited to operating temperatures of 50°C (122°F) or less.

Grade G-10, glass fabric with epoxy resin binder, has extremely high mechanical strength (flexural, impact, and bonding) at room temperature and good dielectric loss and electric strength properties under both dry and humid conditions.

Grade G-11, glass fabric with heat-resistant epoxy resin binder, has properties similar to those of Grade G-10 at room temperature and, in addition, has high retention of flexural strength at elevated temperatures.

Nylon Cloth Grade with Phenolic Resin Binder. *Grade N-1* has excellent electrical properties under high humidity conditions and good impact strength, but is subject to flow or creep under load, especially at temperatures higher than normal.

Flame-Resistant Grades. *Grade FR-1*, paper-based laminates with a phenolic resin binder, are similar in all properties to Grade XP, but so formulated to have at least a UL94 V-1 classification when tested according to UL94.

Grade FR-2, paper-based laminates with a phenolic resin binder, are similar in all properties to Grade XXXPC, but so formulated to have at least a UL94 V-1 classification when tested according to UL94.

Grade FR-3, paper-based laminates with epoxy resin binder, have higher flexural strength than Grade XXXPC and are formulated to have a flame resistance of at least a UL94 V-1 classification when tested in accordance with UL94. The material has low dielectric loss properties with good stability of electrical properties under conditions of high humidity.

Grade FR-4, continuous-filament glass cloth with an epoxy resin binder, is similar in all properties to Grade G-10, but so formulated to have at least a UL94 V-1 classification when tested according to UL94.

Grade FR-5, continuous filament glass cloth with an epoxy resin binder, is similar in all properties to Grade G-11, but so formulated to have at least a UL94 V-1 classification when tested according to UL94.

Composite-Based Laminates. *Grade CEM-1* are laminates with continuous-filament glass cloth surfaces and a cellulose core, all with a flame-resistant epoxy resin binder. With good punching practice, sheets up to and including 2.4 mm (0.094 in.) in thickness may be punched at temperatures not less than 23°C (73°F). These laminates meet UL94 V-0 when tested in accordance with UL94.

Grade CEM-3 is laminated with continuous-filament glass cloth surfaces and a nonwoven glass core, all with a flame-resistant epoxy resin binder. Property values approach those of FR-4. With good punching practice, sheets up to and including 1.6 mm (0.062 in.) in thickness may be punched at temperatures not less than 23°C (73°F). The grade meets UL94 V-0 when tested in accordance with UL94.

Health and Environmental Concerns

Key resins used in the manufacture of laminates are made with formaldehyde (qv). The A-stage resins are manufactured to have low levels of free formaldehyde, less than one percent, and plant atmospheres as well as individual operators are monitored to be certain they are exposed to levels of formaldehyde that are below OSHA guidelines of 0.75 ppm (14).

In the final product, the formaldehyde has completely reacted to form a very inert thermoset resin. Spontaneous emission of formaldehyde from high pressure laminates is measured at approximately the accepted background level of 0.035 ppm (15). Melamine surfaced laminates are tested and approved for food service equipment by the National Sanitation Foundation (16).

In fires, melamine–phenolic laminates ignite slowly at high temperatures and burn slowly producing smoke that has about the same toxicity as wood smoke (17).

Disposal of laminate scrap resulting from edge trim, sanding dust, and fabrication trim presents other problems. The scrap within the manufacturing plant can be ground up and used as part of the fuel source in boilers with the proper permits and controls. Stack effluents from treating operations can be burned off in boiler feedstock or captured in charcoal filter beds. Fabrication shop scrap goes to landfills where it has not generally been a problem because of its density and inertness.

Although it would be desirable to recycle laminate scrap, this has been difficult because of its thermoset nature. However, a 1993 patent (18) suggested a means whereby scrap consisting of cellulose, thermoset resins, and partially reacted resins can be ground to a powder which is used as a filler in a thermoplastic resin. The filled thermoplastic resin is then used for molding of various articles.

BIBLIOGRAPHY

"Lamination and Laminated Products" in *ECT* 1st ed., Vol. 8, pp. 185–192, by H. W. Narigan and G. E. Vybiral, Panelyte Division, St. Regis Paper Co.; "Laminated and Reinforced Plastics" in *ECT* 2nd ed., Vol. 12, pp. 188–197, by C. S. Grove, Jr., Syracuse University, and D. V. Rosato, Consultant, Plastics World; in *ECT* 3rd ed., Vol. 13, pp. 968–978, by F. J. McGarry, Massachusetts Institute of Technology.

1. S. A. Giddings, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, Vol. 8, 2nd ed., John Wiley & Sons, Inc., New York, 1987, pp. 617–646.
2. N. Cheremisinoff, ed., *Handbook of Ceramics and Composites*, Marcel Dekker, Inc., New York, 1990.
3. T. S. Carswell, *Phenoplasts: Their Structure, Properties and Chemical Technology*, Interscience Publishers, Inc., New York, 1947.
4. *The Chemistry of Melamine Crystals*, Melamine Chemicals, Inc., Donaldson, La., 1990.
5. E. G. Melby and J. M. Castro, *Comprehensive Polymer Science*, Vol. 7, Pergamon Press, Oxford, U.K., 1989, Chapt. 3.
6. R. E. Carpenter, *You're in the Chemistry Business: Fundamentals of Polyester and Other Chemical Resistant Resins*, Ashland Chemical Co., Columbus, Ohio, 1982.
7. S. H. Goodman, ed., *Handbook of Thermoset Plastics*, Noyes Data Corp., Park Ridge, N.J., 1986.
8. J. Casey, ed., *Pulp and Paper Chemistry and Chemical Technology*, Vol. I, 3rd ed., John Wiley & Sons, Inc., New York, 1980.
9. R. J. Keeling, "The Effect of Volatile Loss During Analysis on the Mechanical Spectra of Phenolic Kraft Composites," *Conference Proceedings, Society of Plastics Engineers ANTEC*, Apr. 30–May 3, New Orleans, La., 1984.
10. R. H. Young, P. W. Kopf, and O. Salgado, *TAPPI* **64**(4) Apr. 1981.
11. T. Grentzer and J. Leckenby, *Am. Lab.*, (Jan. 1989).
12. *High Pressure Decorative Laminates*, Standards Publication No. LD3-1991, National Electrical Manufacturers Association (NEMA), Washington, D.C., 1991.
13. *Industrial Laminated Thermosetting Products*, Standards Publication No. LI1-1989, NEMA, Washington, D.C., 1989.
14. "Occupational Exposure to Formaldehyde, Final Rule," *Fed. Reg.*, Part 11, Dept. of Labor, Occupational Safety and Health Administration, 29 CFR, Part 1910, May 27, 1992.
15. National Particleboard Association, *Large Chamber Test (FTM2-1985) OSHA 52 protocol*, private communication, June 27, 1988.
16. *Standard 35 for Laminated Plastics for Surfacing Food Service Equipment*, National Sanitation Foundation, Ann Arbor, Mich., Nov. 1979.
17. U.S. National Bureau of Standards, *Further Development of a Test Method for the Assessment of the Acute Inhalation Toxicity of Combustion Products*, NBSIR 82-2532, Washington, D.C., June 1982.
18. **U.S. Pat. 5,215,695** (June 1, 1993), C. Bortoluzzi and R. Bogana (to Abet Laminati SpA).

Ronald J. Keeling
Formica Corporation

LAMINATED WOOD-BASED COMPOSITES.

See WOOD BASED COMPOSITES AND LAMINATES.

LAMPBLACK.

See CARBON, CARBON BLACK.

LANOLIN.

See WOOL.

LANTHANIDES

Lanthanides is the name given collectively to the fifteen elements, also called the 4f elements, ranging from lanthanum, La, atomic number 57, to lutetium, Lu, atomic number 71. The rare earths comprise lanthanides, yttrium, Y, atomic number 39, and scandium, Sc, atomic number 21. The most abundant member of the rare earths is cerium, Ce, atomic number 58 (see CERIUM AND CERIUM COMPOUNDS).

A history of the rare earths can be found in Reference 1. The first rare earth was discovered during the investigation of the mineral gadolinite. An impure oxide was isolated by Gadolin in 1796 and called yttria. In 1803, another oxide was reported and named ceria. Sir Humphrey Davey demonstrated in 1808 that earths as a class were not elements, but compounds composed of oxygen and a metallic element. Between 1839 and 1901, the rare earths were extensively studied and all were isolated through extensive fractionation processes. Industrial use started in the beginning of the twentieth century, mainly from the works of Auer von Welsbach on incandescent mantles and lighter flints.

Occurrence

The lanthanides, distributed widely in low concentrations throughout the earth's crust (2), are found as mixtures in many massive rock formations, eg, basalts, granites, gneisses, shales, and silicate rocks, where they are present in quantities of 10–300 ppm. Lanthanides also occur in some 160 discrete minerals, most of them rare, but in which the rare-earth (RE) content, expressed as oxide, can be as high as 60% rare-earth oxide (REO). Lanthanides do not occur in nature in the elemental state and do not occur in minerals as individual elements, but as mixtures.

Usually lanthanides are divided into several subgroups: the light lanthanides, from La to Nd, medium lanthanides, from Sm to Dy, and heavy lanthanides, Ho to Lu. Alternatively, nomenclature such as ceric RE, from La to Nd, and yttric RE, from Sm to Lu plus Y, is used.

The relative abundance of certain rare earths in rocks is a powerful tool for the study of the formation of rock deposits. When the molten basic basalt rocks, which lie deep in the earth, come into contact with the more acidic silicate rocks, the silicates extract some of the rare earths from the basalts. The silicate rocks extract a higher percentage of the light rare earths than of the heavy ones. By studying the relative abundances of these elements in various rocks, geochemists can tell whether or not they have been molten. Most of the rare-earth-rich minerals were formed deep in the earth's crust, precipitating from superheated solutions of seawater or molten rock that were subjected to very high pressures. An abnormally low relative abundance of europium points to a reaction in a reducing atmosphere when the mineral was formed. Conversely, a low abundance of cerium indicates an oxidizing atmosphere. The relative abundances of the rare earths in the earth's crust are listed in Table 1. The abundances of the even-atomic-numbered lanthanides are considerably greater than the adjacent odd-atomic-numbered lanthanides, as is true generally for all elements. The relative abundance of the rare earth in chondritic, ie, stony or granular, meteorites is accepted to be the average distribution of these elements in the original solar system (see EXTRATERRESTRIAL MATERIALS).

Table 1. Rare Earths and Other Elements in the Earth's Crust

RE elementsz	Abundance, ppm	Other elements	Abundance, ppm
lanthanum	18	carbon	320
cerium	46	chromium	200
praseodymium	5.5	manganese	1,000
neodymium	24	iron	50,000
samarium	6.5	cobalt	23
europium	0.5	nickel	80
gadolinium	6.4	copper	70
terbium	0.9	zinc	130
dysprosium	5.0	cadmium	0.3
holmium	1.2	mercury	1
erbium	4.0	silver	0.1
thulium	0.4	gold	0.005
ytterbium	2.7	platinum	0.005
lutetium	0.8	tin	40
scandium	10	lead	16
yttrium	28		

Comparing the relative abundance of the rare earths and the other elements listed in Table 1, the rare earths are not so rare. Cerium, the most abundant of the rare-earth elements is roughly as abundant as tin; thulium, the least abundant, is more common than cadmium or silver. Over 200

rare-earth-containing minerals have been identified. Allantite, apatite, bastnaesite, brannerite, cerite, euxenite, fergusonite, fluocerite, gadolinite, monazite, pyrochlore, samarskite, xenotime , and zircon are considered the principal rare-earth-containing minerals, but only monazite and bastnaesite are processed on a large industrial scale. China has the largest share of the world's reserves of rare earths (more than 75%) . The United States, India, South Africa, and Australia are also significant players in the field (Table 2) (3).

Table 2. World Reserves of Rare Earths

Country	Quantity, t × 10 ³	Distribution, %
United States	6,471	13.6
Australia	754	1.6
India	1,939	4.1
South Africa	987	2.1
China	36,000	75.6
other	1,463	3.1
Total	47,614	100

Properties

In the sixth row of the Periodic Table, the binding energies of the 4*f* and 5*d* subshells lie close together. The electronic configuration of the elements vary from one to the other as a function of only small factors such as the filling of the 4*f* subshell. All the lanthanide elements have [Xe]6*s*²4*f*^{*n*+1} configuration except lanthanum, gadolinium, and lutetium, which have [Xe]6*s*²5*d*¹4*f*^{*n*} (if *n* = 0 for lanthanum). Of most direct interest, however, is the electronic configuration of the ions, which are generally trivalent and have configuration [Xe]4*f*^{*n*}, the 4*f* subshell being progressively filled from lanthanum to lutetium.

The chemical properties of the trivalent lanthanides vary little from one lanthanide to another in a given compound or solution. Whereas the filling of the 4*f* subshell plays a minor role in the chemical properties of the atom it is of primary importance to such physical properties as optical and magnetic properties of rare-earth compounds. Energetically, the 4*f* subshell lies inside the already filled 5*s*² and 5*p* subshells, and is thus shielded from interactions with neighboring atoms. This accounts for the original location of the lanthanides in Mendeleev's table (4) and for very specific properties such as the highly ionic character of metal–ligand bonding or the narrow bandwidths for the *f*–*f* electronic transitions.

Significant differences thus characterize the rare earths when these are compared, for example, to the transition-metal ions. As a result of the very strong spin-orbit coupling in the rare earths, each electronic state is described by a total angular momentum *J* and not through *L* and *S* quantum numbers. The effect of the crystal field induced by the ligand on the rare-earth ion is much smaller than the one of interelectronic repulsions (see COORDINATION COMPOUNDS). A description of the relative energies of the *J* states is given in Reference 5. A consequence of the electronic structure is the lanthanide contraction. As the atomic number increases across the lanthanide series, the increasing charge on the nucleus acts on the electrons and tends to pull them closer to the nucleus, leading to a smooth decrease in ionic radius of the trivalent lanthanide cations from lanthanum to lutetium.

Although rare-earth ions are mostly trivalent, lanthanides can exist in the divalent or tetravalent state when the electronic configuration is close to the stable empty, half-filled, or completely filled shells. Thus samarium, europium, thulium, and ytterbium can exist as divalent cations in certain environments. On the other hand, tetravalent cerium, praseodymium, and terbium are found, even as oxides where trivalent and tetravalent states often coexist. The stabilization of the different valence states for particular rare earths is sometimes used for separation from the other trivalent lanthanides. The chemical properties of the di- and tetravalent ions are significantly different.

Physical Properties. An overview of the metallurgy (qv) and solid-state physics of the rare earths is available (6). The rare earths form alloys with most metals. They can be present interstitially, in solid solutions, or as intermetallic compounds in a second phase. Alloying with other elements can make the rare earths either pyrophoric or corrosion resistant. It is extremely important, when determining physical constants, that the materials are very pure and well characterized. All impurity levels in the sample should be known. Some properties of the lanthanides are listed in Table 3.

The arc and spark spectra of the individual lanthanides are exceedingly complex. Thousands of emission lines are observed. For the trivalent rare-earth ions in solids, the absorption spectra are much better understood. However, the crystal fields of the neighboring atoms remove the degeneracy of some states and several levels exist where only one did before. Many of these crystal field levels exist very close to a base level. As the solid is heated, a number of the lower levels become occupied. Some physical properties of rare-earth metals are thus very sensitive to temperature (7).

Above room temperature, the trivalent lanthanide ions are paramagnetic, with the exception of diamagnetic lanthanum and lutetium. Tetravalent cerium and divalent ytterbium are also diamagnetic. For the metals, when the temperature is lowered, the spin and orbital moments line up. The metals become antiferromagnetic or even ferromagnetic, eg, gadolinium, terbium, or dysprosium. The magnetism of the rare earths is highly anisotropic and is important in some industrial applications.

Chemical Properties. Although the chemical properties of the trivalent lanthanides are quite similar, some differences occur as a consequence of the lanthanide contraction (see Table 3). The chemical properties of yttrium are very similar too, on account of its external electronic structure and ionic radius. Yttrium and the lanthanides are typical hard acids, and bind preferably with hard bases such as oxygen-based ligands. Nevertheless they also bind with soft bases, typically sulfur and nitrogen-based ligands in the absence of hard base ligands.

Table 3. Properties of the Lanthanides^a

Parameter	Lanthanum	Cerium	Praseodymium	Neodymium	Promethium	Samarium	Europium
CAS Registry Number	[7439-91-0]	[7440-45-1]	[7440-10-0]	[7440-00-8]	[7440-12-2]	[7440-19-9]	[7440-53-1]
atomic number	57	58	59	60	61	62	63
atomic weight	138.91	140.12	140.907	144.24	145	150.35	151.96
melting point, °C	918	798	931	1021	1042	1074	822
boiling point, °C	3464	3433	3520	3074	~3000	1794	1429
density, g cm ³	6.1453	6.770	6.77.3	7.007		7.520	5.234
heat of fusion, kJ mol ^b	6.201	5.179	6.912	7.134		8.623	9.221
heat of sublimation, at 25°C, kJ mol ^b	431.0	422.6	355.6	327.6	~348	206.7	144.7
conduction electrons	3	3, 3.1	3	3	3	3	2
crystal structure	hcp	dhcp	dhcp	dhcp		rhomb	bcc
radius of atom, nm	0.1879	0.1824	0.1828	0.1821	0.1811	0.1804	0.20418
Curie point, °C							
Néel point, °C		ca 13				15	90
valence in aqueous solutions	3	3,4	3	3	3	3	3,2

color of oxide ^c	white	off-white ^d	black ^e	blue		cream	white, greenish tinge
color of aqueous solution ^c	colorless	colorless ^d	green	rose	yellow	colorless	colorless
ionic radius, nm	0.1061	0.1034	0.1013	0.0995	0.0979	0.0964	0.0950

Table 3. (<i>Continued</i>)								
Parameter	Gadolinium	Terbium	Dysprosium	Holmium	Erbium	Thulium	Ytterbium	Lutetium
CAS Registry	[7440-54-2]	[7440-27-9]	[7429-91-6]	[7440-60-0]	[7440-52-0]	[7440-30-4]	[7440-64-4]	[7439-94-3]
Number								
atomic number	64	65	66	67	68	69	70	71
atomic weight	157.95	158.9254	162.50	164.930	167.26	168.934	173.04	174.97
melting point, °C	1313	1365	1412	1474	1529	1545	819	1663
boiling point, °C	3273	3230	2567	2700	2868	1950	1196	3402
density, g cm ³	7.9004	8.2294	8.5500	8.7947	9.066	9.3208	6.9654	9.8404
heat of fusion, kJ mol ^b	10.05	10.80	10.782	16.874	19.90	16.84	7.657	18.65
heat of sublimation, at 25°C, kJ mol ^b	397.5	288.7	290.4	300.8	317.10	232.2	152.1	427.6
conduction electrons	3	3	3	3	3	3	2	3
crystal structure	hcp	hcp	hcp	hcp	hcp	hcp	fcc	hcp
radius of atom, nm	0.18013	0.17833	0.17743	0.17661	0.17566	0.17462	0.19392	0.17349
Curie point, °C	292.7	220	86	19	18	32		
Néel point, °C		230	178	133	84	56		
valence in aqueous solutions	3	3	3	3	3	3	3	3
color of oxide ^c	white	brown ^f	yellowish white	yellowish white	pink	white, greenish tint	white	white
color of aqueous solution ^c	colorless	colorless	yellow tint	yellow	pink	white, greenish tint	colorless	colorless
ionic radius, nm	0.0938	0.0923	0.0908	0.0894	0.0881	0.0870	0.0858	0.850

^a All elements are silvery in color.

^b To convert J to cal, divide by 4.184.

^c RE₂O₃ unless otherwise noted.

^d CeO₂.

^e Pr O₁₁.

^f Tb₄O₇.

In aqueous solutions, trivalent lanthanides are very stable whereas only a limited number of lanthanides exhibit a stable divalent or tetravalent state. Practically, only Ce⁴⁺ and Eu²⁺ exist in aqueous solutions. The properties of these cations are very different from the properties of the trivalent lanthanides. For example, Ce⁴⁺ is more acidic and cerium(IV) hydroxide precipitates at pH 1. Eu²⁺ is less acidic and europium(II) hydroxide does not precipitate at pH 7–8.5, whereas trivalent lanthanide hydroxides do. Some industrial separations are based on these phenomena.

The chlorides, bromides, nitrates, bromates, and perchlorate salts are soluble in water and, when the aqueous solutions evaporate, precipitate as hydrated crystalline salts. The acetates, iodates, and iodides are somewhat less soluble. The sulfates are sparingly soluble and are unique in that they have a negative solubility trend with increasing temperature. The oxides, sulfides, fluorides, carbonates, oxalates, and phosphates are insoluble in water. The oxalate, which is important in the recovery of lanthanides from solutions, can be calcined directly to the oxide. This procedure is used both in analytical and industrial applications.

The lanthanides readily form double salts, eg, Ln₂(SO₄)₃·Na₂SO₄·2H₂O and Ln₂(SO₄)₃·MgSO₄·24H₂O, which were used historically in the fractionation process.

Anhydrous rare-earth salts cannot be prepared by evaporating water. For example, when heated, LnCl·6H₂O converts partially to the oxychloride. The traditional practice has been to heat the lanthanide chloride hydrate slowly while a stream of anhydrous HCl gas is passed over it. Newer methods involve the preparation of anhydrous lanthanide halides through dehydration of lanthanide halide hydrates using HX or NH₄X, where X = halogen (8). Also, many binary anhydrous salts can be prepared by combining the metal directly with a more electronegative element: for example, Cl₂ over Ln gives LnCl₃ and Ln plus S gives Ln₂S₃.

The lanthanides can form hydrides (qv) of any composition up to LnH₃. Lanthanide hydrides can desorb hydrogen reversibly with temperature. Therefore, the lanthanides and some of their alloys are good candidates for hydrogen (qv) storage, of which LaNi₅ is probably the most promising (see HYDROGEN ENERGY).

Many of the binary compounds of the lanthanides, such as oxides, nitrides, and carbides, can exist as nonstoichiometric compounds. These form crystals where some of the anions are missing from the sites the anions normally occupy.

The lanthanides form many compounds with organic ligands. Some of these compounds are water-soluble, others oil-soluble. Water-soluble compounds have been used extensively for rare-earth separation by ion exchange (qv), for example, complexes form with citric acid, ethylenediaminetetraacetic acid (EDTA), and hydroxyethylethylenediaminetriacetic acid (HEEDTA) (see CHELATING AGENTS). The complex formation is pH-dependent. Oil-soluble compounds are used extensively in the industrial separation of rare earths by liquid–liquid extraction. The preferred extractants are carboxylic acids, organophosphorus acids and esters, and tetraalkylammonium salts.

It is easy to reduce anhydrous rare-earth halides to the metal by reaction of more electropositive metals such as calcium, lithium, sodium, potassium, and aluminum. Electrolytic reduction is an alternative in the production of the light lanthanide metals, including didymium, a Nd–Pr mixture. The rare-earth metals have a great affinity for oxygen, sulfur, nitrogen, carbon, silicon, boron, phosphorus, and hydrogen at elevated temperature and remove these elements from most other metals.

Neutron-rich lanthanide isotopes occur in the fission of uranium or plutonium and are separated during the reprocessing of nuclear fuel wastes (see NUCLEAR REACTORS). Lanthanide isotopes can be produced by neutron bombardment, by radioactive decay of neighboring atoms, and by nuclear reactions in accelerators where the rare earths are bombarded with charged particles. The rare-earth content of solid samples can be determined by neutron

bombardment and gamma-ray spectrometry or by x-ray spectrometry.

Mining

A limited number of rare-earth minerals are mined for large-scale rare-earth production: monazite, bastnaesite, loparite [12173-83-0], xenotime [13817-22-6]. In addition, since the 1980s rare-earth-containing clays called ionic ore are mined in China. Table 4 shows the rare-earth composition of typical mineral concentrates.

Table 4. Lanthanide and Yttrium Distribution in Mineral Sources, wt %^{a b},

Rare earth	Bastnaesite	Loparite	Monazite			Xenotime
	California	Russia	E. Australia ^c	W. Australia ^c	India ^d	Malaysia
lanthanum	32.0	27.8	20.2	23.9	23.0	0.50
cerium	49.0	57.1	45.3	46.1	46.0	5.00
praseodymium	4.40	3.7	5.40	5.05	5.50	0.70
neodymium	13.5	8.7	18.3	17.4	20.0	2.20
samarium	0.50	0.91	4.60	2.53	4.00	1.90
europium	0.10	0.13	0.10	0.05		0.20
gadolinium	0.30	0.21	2.00	1.49		4.00
terbium	0.01	0.07	0.20	0.04		1.00
dysprosium	0.03	0.09	1.15	0.69		8.70
holmium	0.01	0.03	0.05	0.05		2.10
erbium	0.01	0.07	0.40	0.21		5.40
thulium	0.02	0.07	trace	0.01		0.90
ytterbium	0.01	0.29	0.20	0.12		6.20
lutetium	0.01	0.05	trace	0.04		0.40
yttrium	0.10	0.14	2.10	2.41		60.8

^a Ref. 3.
^b On a basis of 100% REO.
^c Australian monazite usually contains 4–8% thorium and 0.1–0.3% uranium.
^d Indian monazite contains 8–10% thorium.

Monazite contains typically 50 wt % REO, present as phosphate. The light lanthanide fraction (La, Ce, Pr, Nd) is higher than 90% of the total rare-earth content whereas the other lanthanides make up 5 to 10 wt %. Monazite is a by-product of titanium ore mining in Australia, Brazil, India, Korea, Malaysia, Thailand, South Africa, and the United States and a by-product of iron (qv) ore mining in China. Extensive deposits of beach sands are dredged, subjected to gravimetric and magnetic separation in order to separate ilmenite [12168-52-4], rutile [1317-80-2], zircon [14940-68-2], monazite, and xenotime (see MINERAL RECOVERY AND PROCESSING; SEPARATION, MAGNETIC SEPARATION). The monazite and xenotime content of beach sands is generally very low (<1%, average 0.1%) compared with the ilmenite content (>10%). For this reason, the availability of monazite is greatly dependent on the world demand for ilmenite.

There are extensive deposits of the fluorocarbonate mineral bastnaesite near Mountain Pass in southeastern California. Typical rock analyses show 50 wt % calcite, 25 wt % baryte, 15 wt % bastnaesite, and 10% silica. The ore, after being crushed and ground is upgraded by flotation (qv) to ca 60 wt % rare-earth oxide (REO). It can be upgraded further by leaching with hydrochloric acid to give a 70 wt % REO concentrate. The rare-earth content of this concentrate is as high as 99 wt % in light lanthanides (La to Nd), has almost no heavy rare earths, and ca 0.1% yttrium. Bastnaesite also occurs in China as a by-product of iron ore mining. The world's largest deposit of rare earths is found at Bayan Obo in Inner Mongolia, China, where bastnaesite and monazite co-occur as associated minerals in iron ore (9). The main minerals are based on iron, rare earths, and niobium. Bastnaesite and monazite account for about 70 and 30%, respectively, in the rare-earth fraction of the ore, and the Bayan Obo deposit contains an estimated 36,000,000 t of rare-earth oxides, more than 70% of the estimated world rare-earth reserves.

Xenotime, like monazite, is a rare-earth phosphate. Up to 60% of its rare-earth content is yttria [1314-36-9], Y₂O₃. Xenotime has a higher proportion of heavy rare earths than does monazite. Xenotime occurs with monazite in beach sand deposits. A second source of xenotime is that of cassiterite (tin ore) deposits.

Commercial mining of rare-earth reserves began in the late 1800s. Monazite was the principal rare-earth source up until 1965. Thereafter bastnaesite production exceeded monazite production and as of 1992 bastnaesite, which is the world's principal source of rare earths, constituted 65% of world output of rare-earth minerals (see Table 5). In addition to the conventional ores, there are several other rare-earth resources having a low level of industrial production.

Table 5. Annual World Production of Rare-Earth Ores, t of REO

Country	1990	1991	1992
<i>Monazite</i>			
Australia	7,975	5,000	4,000
Brazil	1,100	1,100	1,100
India	2,400	2,750	2,500
North Korea		50	50
Malaysia	1,925	1,050	1,000
South Africa	1,020	715	700
Sri Lanka	120	110	110
Thailand	358	360	400
United States	2,100	1,500	800
Zaire	94	65	70
<i>subtotal</i>	<i>17,092</i>	<i>12,700</i>	<i>10,730</i>
<i>Bastnaesite</i>			
China	16,480	16,150	16,000
United States	22,714	16,465	20,699
<i>subtotal</i>	<i>39,194</i>	<i>32,615</i>	<i>36,699</i>
<i>Xenotime</i>			
Malaysia	130	130	100
Thailand	27	30	30

<i>subtotal</i>	157	160	130
	<i>Other ores</i>		
^a	8,500	8,500	8,500
<i>World total</i>	64,943	53,975	56,059

^a Mainly loparite.

The uranium ore from Elliot Lake, Canada, contains yttrium and lanthanides (see URANIUM AND URANIUM COMPOUNDS). In the Jiangxi province of the People's Republic of China a large reserve of a rare-earth-containing clay contains over 1,000,000 t of REO. This ore is characterized by having a low cerium content (<5%) but a high content in samarium, europium, terbium, and yttrium compared to the main base REO ores (Table 6).

Table 6. Rare-Earth Oxide Distribution in Mineral and Clay Sources, wt %^a

REO	Bastnaesite	Monazite	Xenotime	Ionic clays	
				Longnan	Xunwu
La ₂ O ₃	27.00	23.35	1.20	2.10	29.84
CeO ₂	50.00	45.69	8.00	1.00	7.18
Pr O ₁₁	5.00	4.16	0.60	1.10	7.41
Nd ₂ O ₃	15.00	15.74	3.50	5.10	30.18
Sm ₂ O ₃	1.10	3.05	2.15	3.20	6.32
Eu ₂ O ₃	0.20	0.10	<0.20	0.30	0.51
Gd ₂ O ₃	0.40	2.03	5.00	5.69	4.21
Tb ₄ O ₇		0.10	1.20	1.13	0.46
Dy ₂ O ₃		1.02	9.10	7.48	1.77
Ho ₂ O ₃		0.10	2.60	1.60	0.27
Er ₂ O ₃	1.00	0.51	5.60	4.26	0.88
Tm ₂ O ₃		0.51	1.30	0.60	0.27
Yb ₂ O ₃		0.51	6.00	3.34	0.62
Lu ₂ O ₃		0.10	1.80	0.47	0.13
Y ₂ O ₃	0.30	3.05	59.30	62.90	10.07

^a On basis of 100% REO.

A large deposit of loparite occurs in the Kola Peninsula, Russia. The production of REO reaches 6500 t yr. Loparite contains over 30% of rare-earth oxides from the cerium group. In addition, loparite contains up to 40% titanium oxide and up to 12% niobium and tantalum oxides. Apatite and other phosphorites constitute a substantial resource of rare earths. The REO content is highly variable and ranges from trace amounts to over 1%. Apatite- [1306-05-4] rich tailings of the iron ore at Mineville, New York, have been considered a potential source of yttrium and lanthanides. Rare-earth-rich apatites are found at the Kola Peninsula, Russia, and the Phalaborwa complex in South Africa. In spite of low REO content apatites could become an important source of rare earths because these are processed in large quantities for the manufacturing of fertilizers (qv).

Processing

Industrial Digestion.

Monazite. The commercial digestion process for monazite uses caustic soda. The phosphate content of the ore is recovered as marketable trisodium phosphate and the rare earths as RE hydroxide (10). The usual industrial practice is to attack finely ground monazite using a 50% sodium hydroxide solution at 150°C or a 70% sodium hydroxide solution at 180°C. The resultant mixed rare-earth and thorium hydroxide cake is dissolved in hydrochloric or nitric acid, then processed to remove thorium and other nonrare-earth elements, and processed to recover the individual rare earths (see THORIUM AND THORIUM COMPOUNDS).

Bastnaesite. The commercial 60% REO concentrate of bastnaesite [68909-13-7] can be upgraded to 90% REO by leaching with hydrochloric acid then calcining. In the Molycorp process the flotation concentrate is heated in air at 620°C to remove CO₂ and oxidize cerium to the tetravalent state. The resulting solid is treated with 30% HCl to yield a marketable cerium concentrate containing 60–70% CeO₂ and to dissolve the other rare-earth elements (11). An alternative process exists, which consists of leaching the concentrate with hydrochloric acid. The rare earths become partially dissolved, and an RE fraction combines with the fluorine from the ore. The mixed rare-earth fluoride residue is then decomposed by treatment using caustic soda, and the resulting rare-earth hydroxides leached with HCl. Bastnaesite can be treated with concentrated caustic soda at ca 200°C to form rare-earth hydroxides, which are then dissolved in acid. The sulfuric acid treatment process consists of digesting the concentrate in concentrated sulfuric acid at 400°C then recovering the rare earths as water-soluble sulfates. Impurities such as iron are removed by neutralization.

Loparite. There are two methods used at various plants in Russia for loparite concentrate processing (12). The chlorination technique is carried out using gaseous chlorine at 800°C in the presence of carbon. The volatile chlorides are then separated from the calcium–sodium–rare-earth fused chloride, and the resultant cake dissolved in water. Alternatively, sulfuric acid digestion may be carried out using 85% sulfuric acid at 150–200°C in the presence of ammonium sulfate. The ensuing product is leached with water, while the double sulfates of the rare earths remain in the residue. The titanium, tantalum, and niobium sulfates transfer into the solution. The residue is converted to rare-earth carbonate, and then dissolved into nitric acid.

Ion-Adsorption Deposits. Ion-adsorption clay deposits result from prolonged *in situ* weathering of REO-rich host rocks, most commonly granitic or volcanic rocks, where erosion has been limited to a low extent. The critical requirements for the formation of such deposits are met in southern China (9). A distinctive feature of the RE distribution pattern in the Chinese ion-adsorption ores is cerium deficiency. Although the reserves of REO in ion-adsorption ores are presently estimated to be only 1,000,000 t, these deposits are very important because of the rare-earth distribution compared with conventional ores (monazite, bastnaesite) (see Table 6). Although ion-adsorption ores are of a much lower grade than conventional ones, the mining and processing are straightforward (9). The ore is mined by open pit methods and leached by dilute aqueous salt solutions, eg, aqueous NaCl or NH₄Cl. Over 90% of the rare-earth content is transferred into solution, then precipitated as oxalates, and converted to RE oxides ready for the markets.

Separation Processes. The product of ore digestion contains the rare earths in the same ratio as that in which they were originally present in the ore, with few exceptions, because of the similarity in chemical properties. The various processes for separating individual rare earth from naturally occurring rare-earth mixtures essentially utilize small differences in acidity resulting from the decrease in ionic radius from lanthanum to lutetium. The acidity differences influence the solubilities of salts, the hydrolysis of cations, and the formation of complex species so as to allow separation by fractional crystallization, fractional precipitation, ion exchange, and solvent extraction. In addition, the existence of tetravalent and divalent species for cerium and europium, respectively, is useful because the chemical behavior of these ions is markedly different from that of the trivalent species.

Selective Oxidation. Cerium, the most abundant lanthanide, can be separated easily after oxidation of Ce(III) to Ce(IV), simplifying the subsequent separation of the less abundant lanthanides. Oxidation occurs when bastnaesite is heated in air at 650°C or when the hydroxides are dried in air

at 120–130°C. Once oxidized, Ce(IV) is separated out either by selective dissolution of trivalent species with dilute acid, or by complete dissolution in concentrated acid followed by selective precipitation of ceric hydroxide [12014-56-1] or solvent extraction of cerium(IV) by tributylphosphate in a nitrate medium. In aqueous solution, the oxidation of cerium(III) [18923-26-7] to cerium(IV) [16065-90-0] is carried out by electrolysis, or by treatment with hydrogen peroxide or sodium hypochlorite. Precipitation of hydrated cerium oxide then occurs when the pH is adjusted to 4.

Selective Reduction. In aqueous solution, europium(III) [22541-18-0] reduction to europium(II) [16910-54-6] is carried out by treatment with amalgams or zinc, or by continuous electrolytic reduction. Photochemical reduction has also been proposed. When reduced to the divalent state, europium exhibits chemical properties similar to the alkaline-earth elements and can be selectively precipitated as a sulfate, for example. This process is highly selective and allows production of high purity europium from low europium content solutions (see CALCIUM COMPOUNDS; STRONTIUM AND STRONTIUM COMPOUNDS).

Fractional Precipitation. A preliminary enrichment of certain lanthanides can be carried out by selective precipitation of the hydroxides or double salts. The lighter lanthanides (La, Ce, Pr, Nd, Sm) do not easily form soluble double sulfates, whereas those of the heavier lanthanides (Ho, Er, Tm, Yb, Lu) and yttrium are soluble. Generally, the use of this method has been confined to crude separation of the rare-earth mixture into three groups: light, medium, and heavy.

Fractional Crystallization. Fractional crystallization, used until the early part of the twentieth century, is uneconomical for processing large quantities of lanthanides. Many recrystallization steps are required to recover high purity products. Several salts and double salts have been used: RE(NO₃)₃·2NH₄NO₃·4H₂O for light lanthanide separation (La, Nd, Pr); 2RE(NO₃)₃·3Mg(NO₃)₂·24H₂O for middle lanthanide separation (Sm, Eu, Gd). Bromates and ethylsulfates have been found useful. Fractional crystallization is particularly slow and tedious for the medium and heavy rare earths.

Ion Exchange. Ion exchange (qv), developed during the second World War, has proven to be effective in the separation of high purity rare earths, but generally involves the processing of very dilute aqueous solutions. In the 1950s, the commercial separation of the rare earths was dominated by ion-exchange methods, but technical and economic limitations have restricted its use in industrial-scale separation processes. In synthetic organic cation-exchange resins, the H⁺ cations readily exchange with other cations in solutions that percolate through a bed of the material. Mixed rare earths in aqueous solution are trivalent cations, and are strongly adsorbed by the resin. The rare earths are recovered by elution using a concentrated solution of a monovalent salt, for example, ammonium chloride. If a complexing agent exhibiting significantly different affinities for the various lanthanides is added to the eluant, then a separation occurs.

This procedure, first used with an ammonium citrate eluant, was a considerable improvement over older fractionation methods for separating the light rare earths in quantities from 100 mg to one kilogram. The citrate elution process has been superseded by band-displacement techniques using aminopolycarboxylate eluants, eg, ethylenediaminetetraacetate (EDTA) and hydroxyethylenediaminetriacetate (HEEDTA) among others. The procedure is explained in Reference 13. A preliminary charge of a foreign ion (called the retarding ion) is adsorbed onto the resin. The foreign ion has a lower affinity for the resin than the rare-earth elements (REE). A charge of rare-earth mixture is then adsorbed as a band on the upper part of the resin column. Then band displacement is carried out using an ion that has a greater affinity for the resin than the REE. A possible candidate is NH₄⁺ because NH₄⁺ has no affinity for the complexing agent, whereas REE have a great affinity. Therefore, the eluant used is often an aqueous ammonium aminopolycarboxylate solution. A band of constant length containing the REE moves down through the column and the concentration of the REE in the band remains constant. The band requires a minimum movement along the column to achieve a constant equilibrium state. Then, the band leaves the column and the eluant is subjected to fractionation in order to recover each individual rare earth.

Since 1970, small amounts of REE have continued to be separated by ion exchange for specific grades, eg, >99.9999% purity. Some developments using newer resins or newer equipment have been proposed, but the main limitation of the ion-exchange process development is the low solubility of RE amino polycarboxylates in aqueous solution. Ion exchange was largely replaced by liquid–liquid extraction during the 1960s for this reason.

Liquid–Liquid Extraction. The liquid–liquid extraction process for the rare-earth separation was discovered by Fischer (14). Extraction of REE using an alcohol, ether, or ketone gives separation factors of up to 1.5. The selectivity of the distribution of two rare-earth elements, RE1 and RE2, between two nonmiscible liquid phases is given by the ratio of the distribution coefficients D1 and D2:

$$F = \frac{D1}{D2} = \frac{(RE1)_o}{(RE1)_a} \frac{D2}{(RE2)_o} \frac{(RE2)_a}{D1}$$

where *o* and *a* represent organic and aqueous phases, respectively, and *F* is the separation factor. For two neighboring trivalent rare earths, *F* ranges between 1 and 5. Thus for an effective separation, many single separation operations have to be repeated, batchwise or continuously.

A fully continuous liquid–liquid extraction process in countercurrent flow enables separation into two groups of REE. Thus the separation of *n*REE requires *n* + 1 countercurrent separations carried out using a column or mixer–settler arrangement (see EXTRACTION, LIQUID–LIQUID). This limitation is, however, compensated for by continuous operation. The three-outlet mixer–settler battery is an improvement, but only two streams contain a high purity REE. A schematic flow diagram of a countercurrent flow extraction plant for the continuous separation of two REE or two groups of REE is shown in Figure 1. The mixture that is to be separated is fed to an intermediate stage of the contactor operating in countercurrent flow. The solvent becomes preferentially charged with the REE that form the most stable complex, while the REE that form the less stable REE complex remain in the aqueous phase. The flow extract is further washed by the scrubbing aqueous solution in order to remove traces of the less stable REE complex. The pure extract is then subjected to back extraction in order to recover a high purity aqueous solution. The solvent is then recycled.

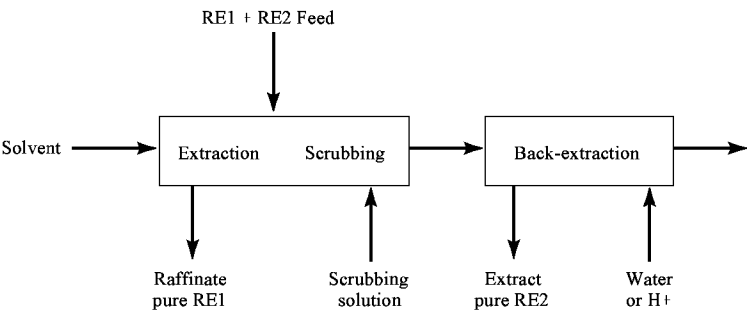


Fig. 1. Schematic flow diagram of a countercurrent extraction unit for the continuous separation of two REE or two groups of REE.

Another characteristic of the solvent extraction system is the high solute concentration in both aqueous and organic phases, which influences greatly the size of the required installation. Concentrations of rare-earth oxides (REO) higher than 100 g/L are often reached in both phases. The process therefore requires only relatively small equipment.

The development of a suitable solvent system is important for successful operation. Solvent systems generally consist of at least the following components: extractant, diluent, inorganic salts or acids, and water. The relative optimization of these components yields the best conditions with which to achieve separation. A key factor to success is the choice of the appropriate extractant. Many extractants may be used for REE separation. These may be divided into three groups on the basis of the mechanisms involved. These extractants are listed in Table 7.

Table 7. Commercial Extractants Available for Rare-Earth Separation

Extractants	CAS Registry Number	Molecular formula	Manufacturer ^a
<i>Acidic extractants</i>			
bis(2-ethylhexyl)phosphoric acid (HDEHP)	[298-07-7]	(C ₈ H ₁₇ O) ₂ POOH	A,D,B,Ho
2-ethylhexyl-2-ethylhexyl-phosphonic acid (HEHEHP)		(C ₈ H ₁₇ O)C ₈ H ₁₇ POOH	A,D,B
bis(2,4,4-trimethylpentyl)-phosphinic acid		(C ₈ H ₁₇) ₂ POOH	AC
neodecanoic acid	[29662-90-6]	C ₉ H ₁₉ COOH	E,S
<i>Basic extractants</i>			
trialkyl methyl ammonium chloride		R ₃ CH ₃ N ⁺ Cl	Ho,He,Sh
<i>Neutral extractants</i>			
tributyl phosphate (TBP)	[126-73-6]	(C ₄ H ₉) ₃ PO	A,D,B,AK
dibutylbutylphosphonate (DBBP)		(C ₄ H ₉) ₂ C ₄ H ₉ PO	A,B
tri- <i>n</i> -octylphosphine oxide (TOPO)	[78-50-2]	(C ₈ H ₁₇) ₃ PO	D,AC

^a Albright and Wilson = A; Daiumli:hachi Chemical = D; Bayer = B; Hoechst = Ho; American Cyanamid = AC; Exxon Chemical = E; Shell Chemical = S; Henkel = He; Sherex = Sh; and AK = AKZO.

Acidic Extractants. Acidic extractants, HL, react with REE according to a cation-exchange reaction:



The extent of extraction or back-extraction is therefore governed by the pH of the aqueous phase.

Extraction by carboxylic acids (qv) is carried out in a neutral or weakly acidic medium. The most widely used carboxylic acid is RR'(CH₃)CCOOH, where R plus R' represents seven carbon atoms. Trade names are Versatic 10 (Shell Chemicals) and Neodecanoic acid (Exxon Chemicals). Carboxylic acids can be used either in chloride or in nitrate media and have a better selectivity for light lanthanides than for heavy lanthanide separation.

A large range of commercial acidic organophosphorus extractants is available: dialkylphosphoric acids, (RO)₂POOH; alkyl alkylphosphonic acids, (RO)R'POOH; and dialkylphosphinic acids, R₂POOH. The most popular alkyl chains are the 2-ethylhexyl and 2,4,4-trimethylpentyl. The efficiency of extraction by the organophosphorus acids decreases when the acidity constant decreases, ie, phosphoric derivatives are stronger than phosphonic derivatives and phosphonic derivatives stronger than phosphinic derivatives, whereas the selectivity with respect to the lanthanide series is adversely affected. A typical plot of the distribution coefficient of REE in the HDEHP perchlorate system is given in Figure 2. Both HDEHP and HEHEHP are widely used by rare-earth processors in France, Japan, the United States, and China.

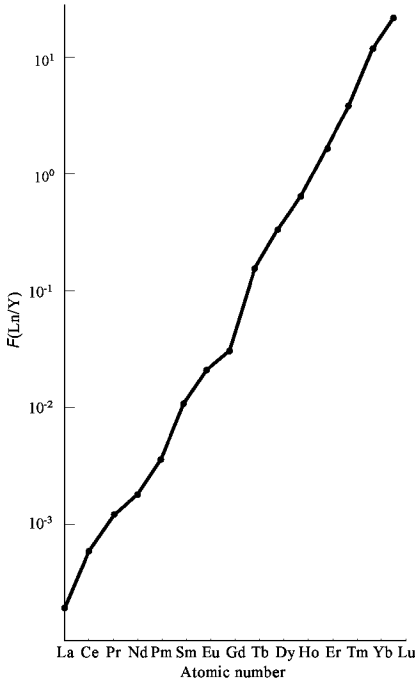
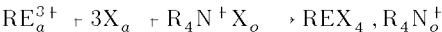


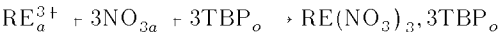
Fig. 2. Distribution of REE in the HDEHP-perchlorate system. The lanthanide–yttrium separation factor, *F*(Ln Y), is plotted as a function of lanthanide atomic number.

Basic Extractants. Only long-chain quaternary ammonium salts, R₃NCH₃⁺ X[−], in which R represents C₈–C₁₂ groups and X nitrate or thiocyanate, are effectively used for REE separations (see QUATERNARY AMMONIUM COMPOUNDS). The extractant reacts with REE according to an anion-exchange reaction:



Therefore the extent of extraction or back-extraction is governed by the concentration of X[−] in the aqueous phase, the distribution coefficients, and selectivities depending on the anion. In nitrate solutions, the distribution coefficient decreases as the atomic number of the REE increases, whereas in thiocyanate solutions, the distribution coefficient roughly increases as the atomic number of the REE increases. The position of yttrium in the lanthanide series is not the same in nitrate and thiocyanate solutions, and this phenomenon has been used for high purity yttrium manufacture in the past. A combination of extraction by carboxylic acids then by ammonium salts is also utilized for production of high purity yttrium.

Neutral Extractants. Many neutral organophosphorus extractants are available: phosphate esters, phosphonate esters, phosphinate esters, and phosphine oxides. The most popular neutral extractant is tributylphosphate (TBP), which reacts with RE elements according to a solvation mechanism:



The extent of extraction can be increased by a salting out effect. The selectivity of TBP is very poor compared with HDEHP and it is only useful for light rare-earth separation; however, organic phase loadings or REO higher than 100 g/L can easily be achieved. There are a large number of TBP manufacturers in Japan, the United States, and Europe.

Equipment. The preferred extraction technique in the rare-earth industry uses mixer-settlers. There are two basic reasons. The first is the use of relatively

small flow rates in the REE refineries. The second results from the large number of stages required for the production of high purity REE, some difficult separations requiring 60 stages and more alone, with some refineries requiring more than 1000 stages.

Recovery. The final step in the chemical processing of rare earths depends on the intended use of the product. Rare-earth chlorides, usually electrolytically reduced to the metallic form for use in metallurgy, are obtained by crystallization of aqueous chloride solutions. Rare-earth fluorides, used for electrolytic or metallothermic reduction, are obtained by precipitation with hydrofluoric acid. Rare-earth oxides are obtained by firing hydroxides, carbonates or oxalates, first precipitated from the aqueous solution, at 900°C.

Metal Production

The deposition of RE metals from aqueous solutions does not work because of the highly electropositive nature of the REE. Therefore, industrial production of RE metals is carried out by fused salt electrolysis or metallothermic reduction.

Fused Salt Electrolysis. Only light RE metals (La to Nd) can be produced by molten salt electrolysis because these have a relatively low melting point compared to those of medium and heavy RE metals. Deposition of an alloy with another metal, Zn for example, is an alternative. The feed is a mixture of anhydrous RE chlorides and fluorides. The materials from which the electrolysis cell is constructed are of great importance because of the high reactivity of the rare-earth metals. Molybdenum, tungsten, tantalum, or alternatively iron with ceramic or graphite linings are used as crucible materials. Carbon is frequently used as an anode material.

Metallothermic Reduction. Rare-earth metals from Nd to Lu, Y, and alloys can be produced by metallothermic reduction. Suitable reducing agents are alkali metals, alkaline-earth metals, aluminium, or light lanthanide metals previously obtained by electrolysis. In some cases, the alloying element can be removed by distillation (qv), yielding the pure rare-earth metal. The anhydrous RE chlorides are preferred in the production of light lanthanides from La to Nd by reduction with calcium at temperatures up to 1100°C, whereas the reduction of the fluorides with calcium is preferred for the production of medium to heavy lanthanides from Gd to Lu at temperatures up to 1600°C. Alternatively, Sm, Eu, and Yb can be produced by reduction of the oxides with a light lanthanide metal including La or mischmetal, a mixture of light lanthanide metals, at 1200°C.

Alloys can be produced by the coreduction process, carried out at 1000°C with calcium, from the oxide mixture. For example, samarium oxide [12060-58-1] and cobalt oxide are coreduced to a SmCo₅ [12017-68-4] powder, CaO then being removed.

Metal Purification. Depending on the relative boiling points, purification may be carried out by RE distillation, alloying element distillation, or zone melting.

Toxicity

The lanthanides are considered only slightly toxic in the Hodge-Sterner classification system and are safely handled with ordinary care (15). Inhalation of rare-earth vapors or dust should be avoided, and the skin washed thoroughly if it comes into contact with any dust or solution.

A toxic effect of the rare earths on humans has not been reported, but extensive tests of toxicity have been made on animals. If the rare earths are administered orally, the toxicity is low. When RE vapors or dusts are inhaled, they are somewhat more toxic and are only slowly absorbed into the body. If injected subcutaneously, most of the injected material remains in place. The most toxic reactions are obtained if the RE are introduced by means of intraperitoneal or intravenous injections. The symptoms of toxicity of the rare-earth elements include writhing, ataxia, labored respiration, walking on the toes with arched back, and sedation. There is a delayed lethality with the death rate peaking between 48 and 96 hours. Chelating agents (qv), citrate, or EDTA, obscure the lethal effects of the rare earths. The effect of atomic weight of rare-earth elements on lethality is difficult to assess, but the medium RE elements appear to have a lesser toxicity than light or heavy RE elements. The toxicity of neodymium salts increases as follows: chloride <propionate <acetate <sulfate <nitrate (15). Some acute lethal dose of lanthanide chlorides by oral or intraperitoneal administration route to mouse are given in Table 8.

Table 8. Acute Lethal Doses of Lanthanide Chlorides for Mice

Element	LD ₅₀ , mg kg	Administration route ^a
La	370	ip
Ce	350	ip
Pr	360	ip
	4500	oral
Nd	600	ip
	5250	oral
Sm	585	ip
	>2000	oral
Eu	550	ip
Gd	550	ip
	>2000	oral
Tb	550	ip
	5100	oral
Dy	585	ip
	7650	oral
Ho	585	ip
	7200	oral
Er	535	ip
	6200	oral
Tm	485	ip
	6250	oral
Yb	395	ip
	6700	oral
Lu	315	ip
	6700	oral

^a ip intraperitoneal.

Uses

Improvements in separation techniques, quality control, and availability of rare-earth compounds in various chemical forms, ie, mixed oxides, metals, and alloys of various purity, morphology, and reactivity, have made these materials an essential part of everyday life.

The extremely unusual physical properties of the rare earths are the reason for a number of industrial applications where no other element can suffice. Furthermore, although RE chemical properties are rather similar to those of the alkaline earths, some specific properties have pushed the rare earths into large industrial developments.

At the beginning of the twentieth century, the incandescent mantle, utilizing the candoluminescence of a mixture of thorium (95% weight) and cerium oxides was developed. The pyrophoricity of rare-earth metals led to the invention of the lighter flint made through the alloying of iron and mischmetal. Since that time, numerous other applications have developed to coincide with the availability of the rare-earth compounds on an industrial scale and having a controlled purity.

Applications Linked to Chemical and Structural Properties.

Metallurgy. The strong affinity for oxygen and sulfur makes the rare-earth metals useful in metallurgy (qv). Mischmetal acts as a trap for these Group 16 (VIA) elements, which are usually detrimental to the properties of steel (qv) or cast iron (qv). Resistance to high temperature oxidation and thermomechanical properties of several metals and alloys are thus significantly improved by the addition of small amounts of mischmetal or its silicide (16,17).

Catalysis. The use of rare earths is mentioned in a number of catalytic reactions (18), but only two areas are considered as industrial applications. One is the stabilization of the zeolites used as catalysts in petroleum (qv) cracking applications, where rare earths (added as chlorides) allow the catalyst to keep a strong acidity, even in the harsh medium in which it is involved (see MOLECULAR SIEVES). Acidity is in fact essential for industrially exploitable conversion of high weight molecules into useful lighter species (19). The second usage is in automotive post-combustion, where cerium oxide is a component of the three-way catalysts planned for usage in all cars before the year 2000. These catalytic systems lower the level of pollutant emissions from cars through selective reduction of nitrogen oxides (NO_x) into nitrogen and water, and simultaneous oxidation of unburnt carbon monoxide and hydrocarbons into carbon dioxide and water vapor. Owing to redox properties, CeO₂ acts as an oxygen reservoir to ensure the buffering effect necessary to control the composition of the exhaust gas, particularly to allow the oxidation of CO and hydrocarbons when the medium is globally reducing. Catalysts are made of precious metal (100–3000 ppm Pd, Rh, or Pt) dispersed on alumina to which 20 wt % of cerium oxide is added. Besides being a buffering agent, and owing to its high thermal stability at the elevated (>800 °C) temperatures in the catalytic muffler, cerium oxide allows alumina to keep a strong surface stability at these temperatures in the catalytic muffler, and gives the metallic particles a good dispersion, avoiding sintering which would make them inactive (20,21). Special grades of cerium oxides, with thermally stable high surface area are developed for this application (see EXHAUST CONTROL, AUTOMOTIVE).

Other catalytic uses of rare-earth compounds have not reached the same development. Neodymium salts are, however, used for rubber manufacturing (22). Divalent samarium halides are employed in organic synthesis (23).

Glass. Rare earths are widely used in the glass (qv) industry, where several properties are advantageous (24). Tetravalent cerium is used to decolorize glass through the oxidation of some of the impurities. For instance, cerium oxidizes iron from its divalent state to the trivalent, changing its color from deep blue to pale yellow. On the other hand, cerium oxide is the best polishing agent known for glass, because of the combination of natural hardness, owing to the cerium oxide compact structure, and a chemical reaction at the silica–cerium oxide interface. Purity and morphology of the polishing powders made from cerium oxide can be adapted to the polishing quality required (25).

Ceramics. Chemical and structural properties of the rare earths are used in the technical ceramics (qv) industry. Tetragonal or cubic forms of zirconia are stabilized by the addition of small quantities of rare-earth oxides, particularly yttrium oxide (1 to 10 mol %). Advantage can be taken of the ionic conductivity induced by the substitution of tetravalent zirconium by trivalent cerium and its variation with oxygen partial pressure to make zirconia-based sensors (qv). The good thermomechanical properties of yttria-stabilized zirconia make it useful for the preparation of cutting tools (see TOOL MATERIALS). Also, when the cubic form of zirconia is fully stabilized, ie, Y₂O₃ content about 7 mol %, optical properties are so close to those of diamond that zirconia is often used in imitation jewelry (26) (see GEMSTONES, GEMSTONE MATERIALS).

Applications Linked to Physical Properties. Applications involving physical properties use high purity (>99.99%) lanthanides and exploit the elements' specific electronic configuration.

Optical Properties. The electronic levels for the 4f electrons of trivalent rare-earths ions are quite insensitive to the crystal field effect and thus have discrete energies. Narrow band electronic transitions are observed in the visible part of the spectrum, resulting in a strong monochromatic character for light absorption or emission, which is of great importance for optical applications (27). Transitions can also occur between 5d band and 4f levels (Eu²⁺, Ce³⁺), or from charge-transfer bands between tetravalent rare earths and anions. Transitions are sensitive to the nature of both the ligand and the cation, and the color can then vary depending on the nature of the crystalline matrix and its composition.

Rare-earth ions are used to give characteristic color to glass or ceramics (see COLORANTS FOR CERAMICS), eg, praseodymium green, neodymium purple, or erbium pink. Tetravalent rare earths are also used. A charge-transfer band between oxygen and praseodymium(IV) in zircon, ZrSiO₄, gives the strongest and most stable yellow pigment for ceramics having high firing temperatures (28). Other optical applications involve the use of cerium(IV) as an antibrowning agent for glass, in TV face plates for example, or of lanthanum as a component (40% by weight) of high index borate glasses for microscopes, telescopes, and camera lenses.

Applications of rare-earth luminescence developed in the early 1960s, as these elements became available in industrial quantities at a high level of purity. Intense and quasi monochromatic emissions obtained from rare-earth activators diluted in appropriate matrices have been used in many applications since that time. Color television was among the first to use rare-earth-based phosphors (see LUMINESCENT MATERIALS, PHOSPHORS). The use of europium-activated yttrium oxysulfide, Y₂O₂S:Eu³⁺, as the red component allowed a twofold increase in brightness compared to ZnS:Ag. Although more expensive, Y₂O₂S:Eu³⁺ has totally superseded ZnS:Ag in commercial television sets (29). The outstanding performance of rare-earth-based phosphors is also utilized in a wide number of professional applications for cathode ray tubes, like monitors for computers, instrument panels in airplanes, or TV projection (30).

Rare-earth-based phosphors are also widely used in tricomponent fluorescent lighting, where light is obtained from the combination of primary monochromatic emissions at 450, 550, and 610 nm. Phosphors combination generally utilize the blue emission of divalent europium, a narrow band from transitions involving 5d levels; green emission from trivalent terbium; and red emission from trivalent europium. Typical phosphors are Sr₅(PO₄)₃Cl:Eu²⁺ and BaMg₂Al₁₆O₂₇:Eu²⁺ for the blue, CeMgAl₁₁O₁₉:Tb³⁺, (Ce, Gd)MgB₅O₁₀:Tb³⁺ and (La,Ce)PO₄:Tb³⁺ for the green, and Y₂O₃:Eu³⁺ for the red. Compact fluorescent lamps are produced using these efficient phosphors that have an energy consumption five to eight times lower than that of incandescent lamps, yet the same color performance (31).

There is also use of rare-earth-based phosphors in x-ray intensifying screens used in medical radiography, where CaWO₄ was historically involved (see MEDICAL IMAGING TECHNOLOGY). The lanthanides have a higher absorption power at the energies utilized, better conversion efficiency (10 to 20% yield compared to 6% for CaWO₄) of incident rays, and the emission of the terbium or thulium activators generally used is perfectly suited to the sensitivity of the film emulsions. Three principal rare-earth-based phosphors are used: Gd₂O₂S:Tb for green, LaOBr:Tm and YTaO₄:Nb or Tm for blues, which allow a significant improvement of the picture quality, and at the same time a lowering exposure time for the patients by a factor of 2 to 4 (32).

Lanthanide luminescence applications have already reached industrial levels of consumption. Additionally, the strongly specific nature of the rare-earths' energy emissions has also led to extensive work in several areas such as photostimulable phosphors, lasers (qv), dosimetry, and fluorescent immunoassay (qv) (33).

Magnetic Properties. Rare-earth metals have exceptional magnetic properties at low temperature. Magnetocrystalline anisotropy constants are 10 to 100 times higher than for other elements, and absolute saturation magnetization is much higher than in iron, for example. However, magnetic ordering occurs only at low temperature, as the internal character of the 4f orbital induces weak couplings for direct interactions between neighboring atoms as well as for long-range exchange by conduction electrons. Thus rare-earth metals are paramagnetic or diamagnetic at room temperature, the highest Curie temperature T_C being 293 K for gadolinium, below which it becomes ferromagnetic (see MAGNETIC MATERIALS, BULK).

In order to raise Curie temperatures, rare-earth metals are alloyed with elements having higher ordering temperatures such as the transition metals iron, cobalt, or nickel. High energy magnets can thus be made. First industrialized were samarium–cobalt magnets (SmCo₅ or Sm₂Co₁₇ based), having T_C above 700°C, energy products of the order of 160 kJ/m³ (20 × 10⁶ GOe), and coercive fields, H_c, around 800 kA/m (10 kOe) (34). Traditional ferrites (qv) and Alnico have values of 32 kJ/m³ (4 × 10⁶ GOe) and 320 kA/m³ (4 kOe), respectively. Such performances allowed intense magnetic energies in small

volumes to be obtained, and thus a miniaturization of devices like stepping motors, or more spectacularly of the small headphones that permitted the appearance of the Walkman in the early 1980s.

Even more powerful are the neodymium–iron–boron magnets that appeared in the mid-1980s (35,36). Magnets derived from the Nd₂Fe₁₄B structure exhibit the highest performances obtained industrially: energy products are >320 kJ/m³ (>40 × 10⁶ GOe) and coercive fields around 950 kA/m³ (12 kOe). These magnets are in the course of a rapid development, eg, in voice coil motors and the automotive industry. There is a promising forecast for several industrial applications.

A relatively new field for the application of rare-earth–transition–element alloys is magnetooptical recording. The magnetic (high coercive field and low*T*_C) and optical (high value of the Kerr rotation angle) properties of (Gd, Tb)(Co, Fe) amorphous alloys are used to obtain high (20 Mbit/cm²) recording densities, in erasable–rewritable laser systems like the minidisk (see INFORMATION STORAGE MATERIALS) (37).

Electrical and Nuclear Properties. Rare earths are often used as additives to modify the performance of several electronic components. Neodymium stabilizes the thermal variation of the dielectric constant of barium titanate, BaTiO₃, over a large temperature range (NPO-type capacitors). Doping by trivalent rare earths gives BaTiO₃ semiconducting properties useful for applications involving electromagnetic wave absorption or to obtain the positive temperature coefficient (PTC) effects used in the realization of sensors or thermal relays. The discovery of one of the most famous high *T*_C superconductors that works using liquid nitrogen involving rare earths is YBa₂Cu₃O_{7.δ} (38) which has a critical temperature over 90 K. These materials are not, however, ready for industrialization.

Gadolinium's extremely high cross section for thermal neutrons, 4.6 × 10^{−24} m² (46,000 barns) per atom, is the reason for its extensive use in the nuclear energy (see NUCLEAR REACTORS). It is used as a component of the fuel or control rods, where it acts as a consumable poison, a trap for neutrons in the reactor (39).

Future Applications. The use of gadolinium complexes as contrast agents in magnetic resonance imaging (mri) is growing (40). Lanthanum–nickel-based alloys are good candidates for rechargeable batteries (qv) technologies (41). Furthermore, although these are not on an industrial scale of usage as of this writing (ca 1994), cerium additives are becoming important for control of pollution (soots, NO_x), from diesel engines. Perovskite-mixed oxides such as LaTO₃, T = Co, Ni, Cu, ...) are being developed for oxidation catalysis (42) and electrodes or interconnects for solid oxide fuel cells (qv) (43). Cerium sulfide is to be used as an alternative to cadmium sulfoselenide reds in the pigment industry (see PIGMENTS, INORGANIC) (44).

Economic Aspects

In 1990 world consumption of lanthanides was approximately 35,000 metric tons (45). The most important markets were the United States/Canada (32.8%), China (18.6%), Europe (15.8%), Japan (14.5%), Eastern Europe (9.5%), the rest of Asia (7.3%), and the rest of the world (1.4%). The principal rare-earth manufacturers in 1993 were Molycorp Inc. and Rhône-Poulenc in the United States; Rhône-Poulenc and Treibacher Chemische WAG in Europe; Shinetsu Chemical, Nippon Yttrium, Mitsubishi Chemical Inc., and Santoku Metal Inc. in Japan; Indian Rare Earths in India; and several additional companies located in the CIS and in the Baotou, Gansu, Yue Long, and Jiangxi provinces in China.

During the period 1990–2000, the expected average annual growth in overall rare-earth compounds demand is from 3–6% in tonnage. In 2000, the expected world consumption might be roughly 56,000 t, and Asia is expected to account for 50% of the total consumption, the largest market being China (25%). During the 1990s the usage of nonseparated rare earths is expected to increase less than usage of separated materials. The latter should double from 8000 t in 1990 to 16,000 t in 2000. In the market of separated rare earths, magnetism is expected to drive the demand to 8000 t in 2000 as compared to 2000 t in 1990. The other markets including phosphors, metallurgy, catalysts, glass, and ceramics are expected to increase at approximately 5% growth annually. Prices of the rare-earth oxides vary considerably from one another depending on their occurrence, and on the specificity required (purity, morphology, etc) for the application. Typical range is between \$20 (light rare-earth oxides) and \$1000 (4*N* Eu₂O₃).

BIBLIOGRAPHY

"Rare-Earth Metals" in *ECT* 1st ed., Vol. 11, pp. 503–521, by H. E. Kremers, Lindsay Chemical Co.; "Rare-Earth Elements" in *ECT* 2nd ed., Vol. 17, pp. 143–168, by W. L. Silvernail and N. J. Goetzinger, American Potash & Chemical Corp.; in *ECT* 3rd ed., Vol. 19, pp. 833–854, by F. H. Spedding, Iowa State University.

1. F. R. Morral, *CIM Bull.* **83**, 97 (1990).
2. G. Muecke and P. Moller, *Sci. Am.* **258**, 72 (1988).
3. J. B. Hedrick, *Ceram. Bull.* **67**, 858 (1988).
4. C. A. Hampel and G. G. Hawley, eds., *Encyclopedia of Chemistry*, 3rd ed., Van Nostrand Reinhold Co., New York, 1973.
5. J. P. Hessler and W. T. Camall, *ACS Symp. Ser.* **131**, 349 (1980).
6. K. A. Gschneider, *J. Less. Comm. Metals* **100**, 1 (1984).
7. R. J. Elliott, ed., *Magnetic Properties of the Rare Earth Metals*, Plenum Press, London, 1972.
8. M. D. Taylor, *Chem. Rev.* **62**, 503 (1962).
9. A. L. Clark and S. Zheng, *Mat. Sci. For.* **8**, 577 (1991).
10. **U.S. Pat. 2,783,125** (1957), C. de Rohden and M. Peltier.
11. **U.S. Pat. 2,722,471** (1955), M. Hirsh, R. F. Bacon, and E. A. Lucas.
12. V. D. Kosynkin and co-workers, *J. All. Comp.* **192**, 118 (1993).
13. J. E. Powell, F. H. Spedding, and D. B. James, *J. Chem. Ed.* **37**, 629 (1960).
14. W. Fischer, W. Dietz, and O. Jubernann, *Naturwiss.* **25**, 348 (1937).
15. T. J. Haley, in K. A. Gschneider and L. Eyring, eds., *Handbook on the Physics and Chemistry of the Rare Earths*, Vol. 4, North Holland Publ. Co, Amsterdam, the Netherlands, 1979, p. 40.
16. H. F. Linebarger and T. K. McCluhan, *ACS Symp. Ser.* **164**, 20 (1981).
17. L. A. Luyckx, *ACS Symp. Ser.* **164**, 43 (1981).
18. T. Inui, *J. All. Comp.* **193**(1–2), 47 (1993).
19. D. N. Wallace, *ACS Symp. Ser.* **164**, 101 (1981).
20. J. G. Nunan and co-workers, *Catalysis and Automotive Pollution Control II*, Elsevier, Amsterdam, the Netherlands, 1992, p. 221.
21. A. F. Diwell and co-workers, in Ref. 20, p. 13.
22. M. Yang, C. Cha, and Z. Chen, *Polym. J.* **22**(10), 919 (1990).
23. H. B. Kagan, J. Collin, and J. L. Namy, *J. All. Comp.* **192**, 191 (1993).
24. J. C. Gadea, *L'industrie Céramique*, 854 (1990).
25. R. V. Horigan, *ACS Symp. Ser.* **164**, 101 (1981).
26. R. Stevens, *Magnesium Elektron*, 1 (1986).
27. G. Blasse, *Mat. Chem. Phys.* **31**, 3 (1992).
28. S. Pajakoff, A. Vendl, and G. Banik, *Interam.* **29**(4), 488 (1980).
29. T. Welker, *J. Luminescence* **48**, 48 (1991).
30. T. Oguri, *Kogyo Zairyo* **40**(12), 37 (1992).
31. B. M. J. Smets, *Mat. Chem. Phys.* **16**, 283 (1987).

32. L. H. Brixner, *Mat. Chem. Phys.* **16**, 253 (1987).

33. G. Blasse, *J. All. Comp.* **192**, 17 (1993).

34. K. Strnat and co-workers, *J. Appl. Phys.* **38**(3), 1001 (1967).

35. M. Sagawa and co-workers, *J. Appl. Phys.* **55**(6), 2083 (1984).

36. J. J. Croat and co-workers, *J. Appl. Phys.* **55**(6), 2078 (1984).

37. F. J. A. M. Greidanus and W. Bas Zeper, *Mat. Res. Soc. Bull.* **15**(4), 31 (1990).

38. J. B. Goodenough, *Phase Trans.* **22**, 79 (1990).

39. M. Yvars, *Ann. Chimie* **10**, 197 (1985).

40. R. B. Laufer, *Chem. Rev.* **87**, 901 (1987).

41. T. Sakai and co-workers, *J. All. Comp.* **180**, 37 (1992).

42. T. Seyama, in L. Tejuca and J. Fierro, eds., *Properties and Applications of Perovskite Type Oxides*, Marcel Dekker, New York, 1993.

43. B. C. H. Steele, *Proceedings of the Conference on Ceramics in Energy Applications*, Sheffield, U.K., 1993, p. 173.

44. D. Huguenin and co-workers, *Proceedings of the RETEC Conference*, Orlando, Fla., 1993.

45. P. Falconnet, *J. All. Comp.* **192**, 114 (1993).

General References

Gmelin Handbuch der Anorganische Chemie, System No. 39, Rare Earth Metals, 8th ed., Springer-Verlag, Berlin.

F. Villani, *Rare Earth Technology and Applications*, Noyes Data Corp., Park Ridge, N.J., 1980.

F. H. Spedding and A. H. Daane, *The Rare Earths*, John Wiley & Sons, Inc., New York, 1961.

K. A. Gschneider, Jr. and L. Eyring eds., *Handbook on the Physics and Chemistry of Rare Earths*, Vols. 1–16, Elsevier, Amsterdam, the Netherlands.

"Industrial Applications of Rare Earth Elements," *ACS Symp. Ser.*, 164, 1981.

The Rare Earth Information Center (RIC), Ames Laboratory of the U.S. Dept. of Energy, Ames, Iowa, maintains an up-to-date file on the new rare-earth developments and publishes quarterly RIC news.

Jean-Louis Sabot
Patrick Maestro
Rhône-Poulenc Recherches

LANTHANUM.

See LANTHANIDES.

LARD.

See FATS AND FATTY OILS; MEAT PRODUCTS.

LASER DYES.

See LASERS.